

VIDEO TAPE RECORDER**Publication number:** JP8292790 (A)**Publication date:** 1996-11-05**Inventor(s):** TANAKA TOMOSHI**Applicant(s):** SANYO ELECTRIC CO**Classification:**

- International: H04N5/783; G10L11/02; G10L19/00; G10L21/04; G11B20/02;
 H04N5/93; H04N5/783; G10L11/00; G10L19/00; G10L21/00;
 G11B20/02; H04N5/93; (IPC1-7): G10L3/02; G10L9/00;
 G10L9/18; G11B20/02; H04N5/783; H04N5/93

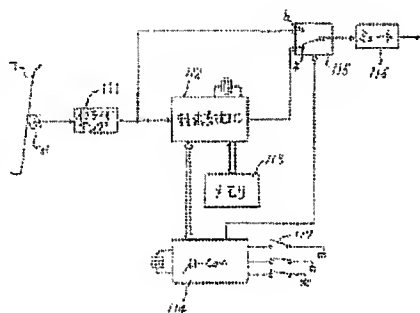
- European:

Application number: JP19950095335 19950420**Priority number(s):** JP19950095335 19950420**Also published as:**

JP3162945 (B2)

Abstract of JP 8292790 (A)

PURPOSE: To perform a speech speed changing so that the deviation between a video and a voice is made smaller and voice information are obtained as much as possible at the time of a double speed mode and to reproduce one part of the voice at a normal speed by a thinning processing at the time of a reproducing equal to or faster than a triple speed mode. **CONSTITUTION:** At the time of a double speed reproducing, this tape recorder has a double speed mode performing the speech speed changing performing a compression or elimination processing with respect to an input voice signal according to whether a reproducing signal is voice sections or silent sections.; Then, at the time of a $\pm N$ (N is a natural number of ≥ 3) multiple speed reproducing, this tape recorder is made so as to be changed over even to an N multiple speed reproducing mode performing the thinning processings of voice sections of prescribed periods of the reproducing voice signal according to the reproducing speed by the command of a system microcomputer 114.



Data supplied from the esp@cenet database — Worldwide

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平8-292790

(43)公開日 平成8年(1996)11月5日

(51)Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L	3/02		G 1 0 L 3/02	A
	9/00		9/00	D
	9/18		9/18	H
G 1 1 B	20/02		G 1 1 B 20/02	G
H 0 4 N	5/783		H 0 4 N 5/783	G

審査請求 未請求 請求項の数5 O L (全 33 頁) 最終頁に続く

(21)出願番号 特願平7-95335

(22)出願日 平成7年(1995)4月20日

(71)出願人 000001889

三洋電機株式会社

大阪府守口市京阪本通2丁目5番5号

(72)発明者 田中 智志

大阪府守口市京阪本通2丁目5番5号 三

洋電機株式会社内

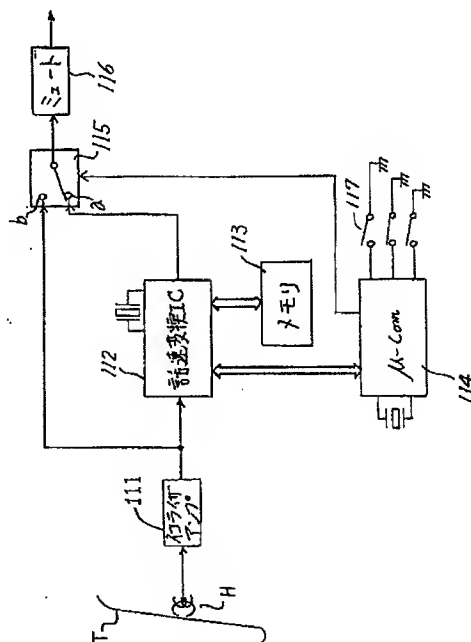
(74)代理人 弁理士 安富 耕二

(54)【発明の名称】 ビデオテープレコーダ

(57)【要約】

【目的】 この発明は、2倍速再生時には映像と音声のズレを小さくでき、しかも可及的に音声情報が得られるよう話速変換を行い、3倍速以上の再生のときには間引き処理によって音声の一部は通常速度で再生できるようにすること。

【構成】 2倍速再生時においては、再生音声信号が音声区間であるか無音区間であるかに応じて、入力音声信号に対して圧縮伸長処理または削除処理を行う話速変換を行う2倍速音声再生モードと、 $\pm N$ 倍速 ($N:3$ 以上の自然数) 再生時においては、再生音声信号の所定の期間の音声区間をこの再生倍速に応じて間引く処理を行う N 倍速再生モードとをシステムマイクロコンピュータ114の指令によって切り替えるようにしたビデオテープレコーダである。



【特許請求の範囲】

【請求項1】2倍速再生時においては、再生音声信号が音声区間であるか無音区間であるかに応じて、入力音声信号に対して圧縮伸長処理または削除処理を行なう話速変換を行う2倍速音声再生モードと、

±N倍速（N：3以上の自然数）再生時においては、再生音声信号の所定の期間の音声区間をこの再生倍速に応じて間引く処理を行うN倍速再生モードと、を設定する制御手段を備えてなるビデオテープレコーダ。

【請求項2】請求項1において、2倍速音声再生モードを行うべく、入力音声信号を話速変換処理する話速変換処理手段、話速変換処理手段の出力が書き込まれるリングメモリ、およびリングメモリからデータを一定速度で読み出す手段を備え、話速変換処理手段は、入力音声信号が音声区間であるか無音区間であるかおよびリングメモリの蓄積量に応じて、入力音声信号に対して圧縮伸長処理または削除処理を行なう手段を備えた話速変換装置を有することを特徴とするビデオテープレコーダ。

【請求項3】請求項1において、2倍速音声再生モードを行うべく、入力されるアナログ音声信号を設定された再生速度倍率に応じたサンプリング周波数でサンプリングするA/D変換手段、A/D変換手段から出力された音声信号が入力されるフレームメモリ、フレームメモリに所要数の音声信号が入力されるごとに、それらの音声信号に対して話速変換処理を行なう話速変換処理手段、話速変換処理手段の出力が書き込まれるリングメモリ、リングメモリから一定速度でデータを読み出す読出手段、およびリングメモリの書き込み信号と読み出し信号とに基づいて、リングメモリの蓄積量を算出する蓄積量算出手段を備えており、話速変換処理手段は、フレームメモリに入力された所要数の音声信号に対応する入力音声信号が、音声区間か無音区間かを判別する区間判別手段、ならびに、区間判別手段の出力および蓄積量算出手段の出力に応じて、上記所要数の音声信号に対して圧縮伸長処理または削除処理を行なう信号処理手段を備えた話速変換装置を有することを特徴とするビデオテープレコーダ。

【請求項4】請求項1において、2倍速音声再生モードを行うべく、入力されるデジタル音声信号が、設定された再生速度倍率に応じた速度で書き込まれるフレームメモリ、フレームメモリに所要数の音声信号が入力されるごとに、それらの音声信号に対して話速変換処理を行なう話速変換処理手段、話速変換処理手段の出力が書き込まれるリングメモリ、1倍速再生時のフレームメモリへの書き込み速度と等しい周波数の読み出し信号に基づいて、リングメモリからデータを読み出す読出手段、およびリングメモリの書き込み信号と読み出し信号とに基づいて、リングメモリの蓄積量を算出する蓄積量算出手段を備えており、

話速変換処理手段は、フレームメモリに入力された所要数の音声信号に対応する入力音声信号が、音声区間か無音区間かを判別する区間判別手段、ならびに、

区間判別手段の出力および蓄積量算出手段の出力に応じて、上記所要数の音声信号に対して圧縮伸長処理または削除処理を行なう信号処理手段を備えている話速変換装置を有することを特徴とするビデオテープレコーダ。

【請求項5】請求項1において、N倍速再生モードを行うべく、N倍速で音声データをメモリに書き込み、1倍速でその書き込まれたデータを読み出すようにメモリを制御するメモリ制御手段を備えたことを特徴とするビデオテープレコーダ。

【発明の詳細な説明】

【0001】

【産業上の利用分野】この発明は、音声信号の話速を変える話速変換装置を備えた、ビデオテープレコーダ（VTR）に関する。

【0002】

【従来の技術】2倍速でも音声を通常のスピードで聞けるようにしたVTRが商品化されており、その基本的構成については例えば雑誌「エレクトロニクス」1993年4月号、34頁～37頁に記載されている。

【0003】

【発明が解決しようとする課題】しかしながら、上記のようなVTRは2倍速再生時においては、通常のスピードで音声聞けるものの、音声情報を無条件に半分間引いて削除してしまう方法であるため、再生音声から記録されたテープの内容を把握することができなくなったり、映像と音声の時間的なズレが大きいという問題点があった。

【0004】従って本発明は2倍速再生時には映像と音声のズレを小さくでき、しかも可及的に音声情報が得られるように音声スピードを制限するように話速変換を行い、2倍速より早い、例えば早送り、巻戻し再生時の速度（5倍、9倍）で、音声速度を再生するときには従来の2倍速再生と同様の方法により音声を再生するようにしたビデオテープレコーダを提供しようとするものである。

【0005】

【課題を解決するための手段】この発明によると2倍速再生時においては、再生音声信号が音声区間であるか無音区間であるかに応じて、入力音声信号に対して圧縮伸長処理または削除処理を行なう話速変換を行う2倍速音声再生モードと、±N倍速（N：3以上の自然数）再生時においては、再生音声信号の所定の期間の音声区間をこの再生倍速に応じて間引く処理を行うN倍速再生モードとを設定する制御手段をビデオテープレコーダに備えてなる。

【0006】また、上記のビデオテープレコーダにおいて、2倍速音声再生モードを行うべく、入力音声信号を

話速変換処理する話速変換処理手段、話速変換処理手段の出力が書き込まれるリングメモリ、およびリングメモリからデータを一定速度で読み出す手段を備え、話速変換処理手段は、入力音声信号が音声区間であるか無音区間であるかおよびリングメモリの蓄積量に応じて、入力音声信号に対して圧縮伸長処理または削除処理を行なう手段を備えた話速変換装置を有することを特徴とする。

【0007】また、上記のビデオテープレコードにおいて、2倍速音声再生モードを行うべく、入力されるアナログ音声信号を設定された再生速度倍率に応じたサンプリング周波数でサンプリングするA/D変換手段、A/D変換手段から出力された音声信号が入力されるフレームメモリ、フレームメモリに所要数の音声信号が入力されるごとに、それらの音声信号に対して話速変換処理を行なう話速変換処理手段、話速変換処理手段の出力が書き込まれるリングメモリ、リングメモリから一定速度でデータを読み出す読出手段、およびリングメモリの書き込み信号と読み出し信号とに基づいて、リングメモリの蓄積量を算出する蓄積量算出手段を備えており、話速変換処理手段は、フレームメモリに入力された所要数の音声信号に対応する入力音声、音声区間か無音区間かを判別する区間判別手段、ならびに、区間判別手段の出力および蓄積量算出手段の出力に応じて、上記所要数の音声信号に対して圧縮伸長処理または削除処理を行なう信号処理手段を備えた話速変換装置を有することを特徴とする。

【0008】また、上記のビデオテープレコードにおいて、2倍速音声再生モードを行うべく、入力されるデジタル音声信号が、設定された再生速度倍率に応じた速度で書き込まれるフレームメモリ、フレームメモリに所要数の音声信号が入力されるごとに、それらの音声信号に対して話速変換処理を行なう話速変換処理手段、話速変換処理手段の出力が書き込まれるリングメモリ、1倍速再生時のフレームメモリへの書き込み速度と等しい周波数の読み出し信号に基づいて、リングメモリからデータを読み出す読出手段、およびリングメモリの書き込み信号と読み出し信号とに基づいて、リングメモリの蓄積量を算出する蓄積量算出手段を備えており、話速変換処理手段は、フレームメモリに入力された所要数の音声信号に対応する入力音声、音声区間か無音区間かを判別する区間判別手段、ならびに、区間判別手段の出力および蓄積量算出手段の出力に応じて、上記所要数の音声信号に対して圧縮伸長処理または削除処理を行なう信号処理手段を備えている話速変換装置を有することを特徴とする。

【0009】上記ビデオテープレコードにおいて、N倍速再生モードを行うべく、N倍速で音声データをメモリに書き込み、1倍速でその書き込まれたデータを読み出すようにメモリを制御するようにしたことを特徴とする。

【0010】

【作用】この発明によれば、2倍速再生時においては、適応型話速変換処理を行い、3倍速以上のときは単純間引き処理を行うようにVTRが制御される。

【0011】この発明によれば、2倍速再生時においては入力音声信号が音声区間であるか無音区間であるかに応じて、入力音声信号に対して圧縮伸長処理または削除処理が行われる。

【0012】この発明によれば、2倍速再生時においては、入力音声信号は話速変換処理手段によって、話速変換処理される。話速変換処理手段の出力は、リングメモリに書き込まれる。リングメモリに書き込まれたデータは、一定速度で読み出される。話速変換処理手段においては、入力音声信号が音声区間であるか無音区間であるかおよびリングメモリの蓄積量に応じて、入力音声信号に対して圧縮伸長処理または削除処理が行なわれる。

【0013】この発明によれば、2倍速再生時においては、入力されるアナログ音声信号は、A/D変換手段により、設定された再生速度倍率に応じたサンプリング周波数でサンプリングされる。A/D変換手段から出力された音声信号は、フレームメモリに入力される。フレームメモリに所要数の音声信号が入力されるごとに、話速変換処理手段により、それらの音声信号に対して話速変換処理が行なわれる。話速変換処理手段の出力は、リングメモリに書き込まれる。リングメモリに書き込まれたデータは、1倍速再生時のサンプリング周波数に等しい周波数の読み出し信号に基づいて読み出される。リングメモリの書き込み信号と読み出し信号とに基づいて、蓄積量算出手段によって、リングメモリの蓄積量が算出される。

【0014】この発明によれば、2倍速再生時においては、フレームメモリに入力された所要数の音声信号に対する入力音声、区間判別手段により、音声区間か無音区間かが判別される。そして、区間判別手段の出力および蓄積量算出手段の出力に応じて、上記所要数の音声信号に対して圧縮伸長処理または削除処理が行なわれる。

【0015】この発明によれば、2倍速再生時においては、入力されるデジタル音声信号が、設定された再生速度倍率に応じた速度でフレームメモリに書き込まれる。フレームメモリに所要数の音声信号が入力されるごとに、話速変換処理手段により、それらの音声信号に対して話速変換処理が行なわれる。話速変換処理手段の出力は、リングメモリに書き込まれる。リングメモリに書き込まれたデータは、読み出し信号に基づいて一定速度で読み出される。リングメモリの書き込み信号と読み出し信号とに基づいて、蓄積量算出手段によって、リングメモリの蓄積量が算出される。

【0016】話速変換処理手段においては、フレームメモリに入力された所要数の音声信号に対する入力音声、区間判別手段により、音声区間か無音区間かが判別

される。そして、区間判別手段の出力および蓄積量算出手段の出力に応じて、上記所要数の音声信号に対して圧縮伸長処理または削除処理が行なわれる。

【0017】本発明によれば、 $\pm N$ 倍速再生時に一部の音声が増長されて、残余の信号が間引かれる

【0018】

【実施例】以下、図面を参照して、この発明の実施例について説明する。

【0019】図23は本発明を実施したVTRの概略ブロック図を示しており、テープTの横トラックから再生されたモノラル音声信号は、音声ヘッドHにてピックアップされイコライザアンプ111に入力される。イコライザアンプ111でイコライズおよびアンプされた音声信号は切り替え回路115の端子bおよび話速変換IC112に供給される。話速変換IC112の出力は前記切り替え回路115の端子aに供給される。前記切り替え回路115はシステムマイコン114からの指令に基づいて端子aまたはbを選択し、その選択した出力をミュータ回路116を介して出力する。すなわち、テープTから出力される音声は端子bから出力される場合、テープ速度比例した速度の音声信号が出力され、端子aからは再生音声信号に圧縮、伸長、削除処理などの処理が施された信号が出力される。話速変換IC112は、システムマイコン114によって制御されるとともに、メモリ（ダイナミックRAM）113と協働して上記の処理を行う。

【0020】次に図23の動作を図24のフローチャートを参照しつつ説明する。

【0021】まず、システムマイコン114は当該VTRが再生モードかどうかの判断を行う（S1）。ステップ1で再生モードであると判断されると次に2倍速再生キー117が押されたかどうかの判断がステップ2（S2）で行われる。この2倍速再生キーが押されると、テープ速度を2倍速にして、映像および音声が増長されるようになっている。そして、音声に関しては2倍速再生キー117が押されると2倍速の適応話速処理（S3）が行われる。

【0022】この適応話速処理について図25を用いて説明すると、通常速度の再生音声を図25（a）のように時間Tで「話速変換を用いたVTRです」なる音声が増長されるものとし、これに2倍速の適応話速処理をかけると同図（b）のように $1/2$ の1の時間（ $T/2$ ）で再生するが、このために通常速度の音声の単語と単語の間の無音部分、すなわち話をしていない部分を削除し、単語の部分は通常速度の音声と近い話速のままで単語をつなげている。このとき、物音などの周囲の雑音のため完全な無音は存在しないが、無音判定のしきい値を音声の状況に応じて適応的に変化させることによって周囲の雑音と目的とする音声を判別している。尚、無音を取り除いただけでは再生時間が $1/2$ にならない場合、

音声の一部を無音区間の長さに応じて適応的に圧縮することにより、再生時間を $1/2$ にしている。そして、再生が終了したとステップ6（S6）で判断されると再生が終了する。

【0023】ステップ2（S2）で2倍速再生キーが押されていない場合、次にステップ4（S4）で $\pm N$ 倍速モードになったかどうかの判断がされる。この $\pm N$ 倍速モードとは例えば早送り再生、または巻戻し再生時がこれに相当する。そして、このステップ4で $\pm N$ 倍速モードでないと判断されると、ステップ6に戻る。

【0024】ステップ4で $\pm N$ 倍速モードである場合、単純間引き処理による圧縮処理を行い（S5）、ステップ6に進む。

【0025】前記単純間引き処理を図26を参照しつつ説明すると、例えば3倍速での再生音声信号の場合、図26（a）で示すようにT1時間で「ビデオテープレコーダに時間差適応話速変換回路を」なる音声信号が増長されているが、この再生信号に対し、単純間引き処理を行うと、同図（b）に示すようにT2期間である「ビデオテ」が増長されてT1時間で再生される。すなわち、3倍速再生の音声信号を $2/3$ を間引いて（間引かれる期間：TD）標準の話速に変換される。換言すれば、 $T1 : T2 = 3 : 1$ になる。

【0026】また同様に、入力信号が5倍速再生の場合、 $4/5$ を間引いて標準話速に変換し（ $T1 : T2 = 5 : 1$ ）、入力信号が9倍速再生の場合、 $8/9$ を間引いて標準話速に変換する（ $T1 : T2 = 9 : 1$ ）。

【0027】また、逆転N倍速再生時（ $-N$ 倍速再生時）に対する単純間引き処理を図5を参照しつつ説明すると、図27に示す通り逆転入力音声の一部を抽出し正方向の通常話速に変換する。すなわち、図27（a）はN倍速の逆転音声信号であり、同図（b）はそれを間引き処理した場合を示している、5倍速逆転再生の場合は、 $4/5$ を間引いて標準話速に変換し（ $T1 : T2 = 5 : 1$ ）、9倍速逆転再生の場合は、 $8/9$ を間引いて標準話速に変換する（ $T1 : T2 = 9 : 1$ ）。尚、図27において、TDは間引かれる期間を示す。

【0028】上記のように2倍速再生時においては、適応型話速変換処理を行い、3倍速以上のときは単純間引き処理を行うようにしているのは、3倍速以上になると適応型話速変換を行っても、音声を削除する量が多くなり、信号処理が簡単な単純間引きとあまり変わらなくなり、複雑な信号処理を必要とする適応話速をする意味が無いばかりでなくかえって聞きづらくなるからである。従って、このように2倍速再生時においては、適応型話速変換処理を行い、3倍速以上のときは単純間引き処理を行うようにすることにより、話速変換用ICとしてその構成が簡単になるばかりでなく、VTRとしての商品価値を向上せしめることができる。

【0029】図1は、上記話速変換IC内の適応型話速

変換を行う部分に相当する話速変換装置の全体的な構成を示している。

【0030】入力音声信号は、ALCアンプ1で増幅された後、A/D変換部2に送られ、例えば12ビットのデジタル信号に変換される。A/D変換部2の標準サンプリング周波数は、たとえば8KHzである。2倍速再生時には、A/D変換部2のサンプリング周波数 f_{sAD} は、16KHzとなる。

【0031】A/D変換部2の出力は、DSP(Digital Signal Processor)4に送られるとともにレベル検出部3にも送られる。レベル検出部3は、A/D変換部2でA/D変換されたデータが変換レンジの最大値となったときに、ALC(automatic level control)信号をALCアンプ1に出力する。これにより、ALCアンプ1のアンプ利得が制御され、A/D変換部2の入力信号が最大レンジを越えないようにされる。つまり、VTRの再生テープ速度が変化するとALCアンプ1の入力信号レベルも変化する。そこで、レベル検出部3の出力に基づいて、アンプ利得を自動調整することにより、A/D変換部2の入力信号が最大レンジを越えないようにしている。

【0032】DSP4は、2フレーム分の音声信号を記憶できる容量のフレームメモリ5およびフレームメモリ5に記憶された音声信号に対してフレーム単位で話速変換処理を行なう話速変換部6とを備えている。1フレームは、ここでは、200個のサンプリングデータから構成されるものとする。

【0033】フレームメモリ5内の前半領域および後半領域のうち、一方の領域に記憶された1フレーム分の音声信号に対して話速変換部6により処理が行なわれると同時に、他方の領域にA/D変換部2からの信号が蓄積される。そして、この他方の領域に1フレーム分の信号が蓄積されると、今度はその領域内のデータに対して話速変換部6により処理が行なわれると同時に、既に処理が行なわれたデータが記憶されていた上記一方の領域にA/D変換部2からの信号が蓄積される。

【0034】話速変換部6から出力されたデータは、書き込みクロックに基づいてリングメモリ7に書き込まれる。リングメモリ7に書き込まれたデータは、読み出しクロックに基づいて、読み出される。リングメモリ7から読み出された信号は、D/A変換部8によってアナログ信号に変換された後、アンプ10で増幅され、音声出力信号として出力される。

【0035】D/A変換部8のサンプリング周波数 f_{sDA} は、8KHzである。また、リングメモリ7の読み出しクロックの周波数も8KHzである。リングメモリ7としては、 $21845 \times 12 \text{ bit}$ のもの、すなわち、21845ワードのものが用いられている。したがって、リングメモリ7にデータを蓄積できる最大時間(入力信号に対する出力時間の最大遅延時間)は、21

$845 \times 1 / 8000 = 2.73$ 秒となる。

【0036】リングメモリ7に対する書き込みクロックは、アップダウンカウンタ9のアップカウント用入力端子(UP)に入力する。リングメモリ7に対する読み出しクロックは、アップダウンカウンタ9のダウンカウント用入力端子(DOWN)に入力する。アップダウンカウンタ9は、入力された書き込みクロックの総数と入力された読み出しクロックの総数との差(リングメモリ7の蓄積量)をカウントし、そのカウント値を15bitのデジタル信号として出力する。アップダウンカウンタ9の出力は、話速変換部6に送られる。

【0037】図2は、話速変換部6の詳細な構成を示している。

【0038】フレームメモリ5から読み出された音声信号は、パワー計算部11に送られ、1フレーム分の音声信号の平均パワー値Pが算出される。この平均パワー値Pは、サンプリングされた1フレーム内の各音声信号の振幅を i_0, i_1, \dots, i_{N-1} (ただし、 $N=200$)とすると、次の数式1によって求められる。

【0039】

【数1】

$$k = N-1$$

$$P = (1/N) \times \sum_{k=0}^{N-1} (i_k)^2$$

$$k = 0$$

【0040】パワー計算部11で求められた平均パワー値Pは、比較部12に送られる。比較部12には、しきい値メモリ13からしきい値 T_h が送られており、平均パワー値Pがしきい値 T_h 以上($P \geq T_h$)か、平均パワー値Pがしきい値 T_h より小さいか($P < T_h$)が判別される。比較部12からは、平均パワー値Pがしきい値 T_h 以上($P \geq T_h$)のときには現フレームが音声区間であることを示す信号が、平均パワー値Pがしきい値 T_h より小さいときには現フレームが無音区間であることを示す信号が、それぞれ出力される。

【0041】しきい値 T_h としては、A/D変換部2の量子化ビット数が12bitのときには、たとえば、212に設定される。なお、次のようにして、しきい値 T_h を変更するようにしてもよい。すなわち、図2に点線で示すように、パワー一定常状態検出およびしきい値更新部14を設ける。パワー一定常状態検出およびしきい値更新部14は、パワー計算部11からの平均パワー値Pが、所定フレーム数(例えば、40フレーム)にわたって一定であったか否かを判別し、一定であったときには(定常状態)、そのときの平均パワー値Pの2倍の値をしきい値メモリ13に書き込み、しきい値 T_h を更新させる。ただし、更新されるしきい値の最大値は、所定値、たとえば214に制限される。このようにすることにより、定常的に発生している雑音が無音区間として取り扱うことができるようになる。

【0042】また、入力信号の音声区間と無音区間とを、次の数式2で示す各フレームの音声信号のパワー累積値 P_a と所与のしきい値とに基づいて判別するようにしてもよい。

【0043】

【数2】

$$\begin{aligned} k &= N-1 \\ P_a &= \sum (i_k) 2 \\ k &= 0 \end{aligned}$$

【0044】比較部12の出力は、条件分岐部15に送られる。条件分岐部15には、リングメモリ蓄積量状態判別部16の出力が入力している。また、条件分岐部15には、パワー計算部11を介してフレームメモリ5からの、音声信号が送られている。さらに、条件分岐部15には、ポーズ継続長設定メモリ17が接続されている。ポーズ継続長設定メモリ17には、無音区間の削除開始点を決定するためのポーズ継続長 T_{del} (無音削除開始点判別値) が設定されている。

【0045】リングメモリ蓄積量状態判別部16は、アップダウンカウンタ9から送られてきた蓄積量に基づいて、リングメモリ7の状態がオーバーフロー直前状態になったこと、およびリングメモリ7の状態がアンダーフロー直前状態になったことを検出する。

【0046】つまり、オーバーフロー検出用データメモリ18にはオーバーフロー検出用データ T_{max} が、アンダーフロー検出用データメモリ19にはアンダーフロー検出用データ T_{min} が、それぞれ記憶されている。オーバーフロー検出用データ T_{max} は、例えば、リングメモリ7の総ワード数 (TOTAL) 21845より200小さい値21645に設定されている。アンダーフロー検出用データ T_{min} は、例えば、200に設定されている。

【0047】そして、アップダウンカウンタ9から送られてきた蓄積量がオーバーフロー検出用データ T_{max} 以上になると、リングメモリ蓄積量状態判別部16からオーバーフロー直前検出信号が出力される。また、アップダウンカウンタ9から送られてきた蓄積量がアンダーフロー検出用データ T_{min} 以下になると、リングメモリ蓄積量状態判別部16からアンダーフロー直前検出信号が出力される。条件分岐部15は、オーバーフロー直前検出信号が入力されているときにはリングメモリ7がオーバーフロー直前状態であると判別し、アンダーフロー直前検出信号が入力されているときにはリングメモリ7がアンダーフロー直前状態であると判別する。

【0048】条件分岐部15は、比較部12から送られてくる音声区間または無音区間の判別信号と、リングメモリ蓄積量状態判別部16から送られてくるリングメモリ状態に関する検出信号と、ポーズ継続長設定メモリ17に設定されているポーズ継続長 T_{del} とに基づい

て、以下の6つのケースに場合分けを行なう。そして、それに応じて、マルチプレクサ20を制御して、音声信号を所定の処理部に送る。

(1) 第1ケース (case 1)

入力信号が音声区間であり、かつリングメモリ7がオーバーフロー直前状態ではないと判別されたときには、第1ケースとなる。

【0049】この場合には、音声信号は、マルチプレクサ20を介して、ピッチ圧縮伸長手段23に送られる。ピッチ圧縮伸長手段23は、バリエابل スピーチ コントロール (VSC) を行なうものであり、再生速度倍率を n とすると、入力信号に対して、圧縮率 $1/n$ より大きな圧縮率で伸長圧縮処理を行なう。ここで用いられる伸長圧縮法としては、例えば、ポインター移動量制御による重複加算法 (Pointer Interval Control Overlap and Add: PICO LA)、TDHS (TimeDomain Harmonic Scaling) 法等がある。ピッチ伸長圧縮手段23で伸長圧縮処理が行なわれた信号は、デマルチプレクサ27を介してリングメモリ7に送られ、書き込みクロックにしたがって、リングメモリ7に書き込まれる。

【0050】VTRの2倍速再生時においては、A/D変換部2のサンプリング周波数 f_{sAD} は16KHZであり、D/A変換部8のサンプリング周波数 f_{sDA} は8KHZである。このため、音程は元に戻されて出力される。

【0051】従来の一般的な時間軸伸長圧縮においては、VTRの2倍速再生時には圧縮率 $1/2$ で、圧縮される。言い換えれば、2ピッチ周期が1ピッチ周期に間引かれる。このため、出力音声は標準音声速度の2倍速となる。つまり、2倍速再生の通常再生では、出力音声は標準音声速度の2倍速となる。ただし、音程は元のままとする。

【0052】これに対し、図2の話し変換部6に設けられた上記ピッチ伸長圧縮手段23では、圧縮率が $1/2$ より大きな値に設定される。ここでは、圧縮率が $2/3$ に設定されているとする。言い換えれば、3ピッチ周期が2ピッチ周期に間引かれる。このため、出力音声は、標準音声速度の $3/2$ 倍速となる。この場合も音程は、元のままである。このように、圧縮率 $2/3$ で圧縮された場合には、圧縮率 $1/2$ の場合に比べて、 $2/3 - 1/2 = 1/6$ だけ、信号が伸長されることになる。この伸長分が、リングメモリ7の蓄積量となる。

【0053】PICO LAを用いて、入力信号を圧縮率 $2/3$ で圧縮する方法について、図3を用いて簡単に説明する。まず、入力信号からピッチ周期が抽出される。抽出されたピッチ周期を T_p とする。波形Aに対しては、1から0へ直線的に向かう重み (重み関数 K_1) がつけられて、波形A' が作成される。波形Bに対しては0から1に向かう重み (重み関数 K_2) がつけられて、波形B' が作成される。

【0054】そして、それらの波形A'およびB'が加え合わされ、長さTpの波形A' * B'が作成される。これらの重みは、波形A' * B'の前後の接続点での連続性を保つためにつけられている。つぎに、ポインターが、圧縮率に基づいて決まる長さである3Tp分だけ移動され、同様な操作が行なわれる。これにより、3つの波形A、B、Cから2つの波形A' * B'およびCが得られる。このようにして、3ピッチ周期分の信号が、2ピッチ周期分の信号に圧縮される。

【0055】ピッチ伸長圧縮手段23による伸長圧縮法としては、図17(a)、(b)に示すように、ピッチ抽出をすることなく、所定長の固定フレーム長Ts単位で伸長圧縮処理を行うようにしてもよい。固定フレーム長Tsは、たとえば入力データの200個分の長さに設定される。図17の例では、3Tsを2Tsにする例を示している。

【0056】図17(a)の方法では、固定フレーム長Tsの波形A、B、Cのうち、波形Aに対しては、1から0へ直線的に向かう重み(重み関数K1)がつけられて、波形A'が作成される。波形Bに対しては0から1に向かう重み(重み関数K2)がつけられて、波形B'が作成される。

【0057】そして、それらの波形A''およびB''が加え合わされ、長さTsの波形A'' * B''が作成される。これらの重みは、波形A'' * B''の前後の接続点での連続性を保つためにつけられている。そして、次の波形Cに対しては、そのまま出力される。これにより、3つの波形A、B、Cから2つの波形A'' * B''およびCが得られる。このようにして、3Ts分の信号が、2Ts分の信号に圧縮される。

【0058】図17(b)の方法では、固定フレーム長Tsの波形A~Cのうちの波形Aには先頭からたとえば20個のデータに0から1へ直線的に向かう重み(重み関数K3)をつけて波形A'を得る。波形Bには181個目~200個目までの入力データに1から0へ直線的に向かう重み(重み関数K4)をつけて波形B'を得る。そして、波形Cを削除する。次の3つの波形D~Fに対しても、同様な処理が行われる。このようにして、3つの波形A~C(またはD~F)からなる信号は、2つの波形A''およびB''(またはD''およびE'')からなる信号に圧縮される。つまり、3Ts分の信号が、2Ts分の信号に圧縮される。

【0059】上記固定フレーム長単位での伸長圧縮処理を用いた場合には、ピッチ周期ごとの伸長圧縮処理を用いた場合に比べて、音質は低下するが、処理量は軽減される。

【0060】なお、この話速変換装置が英語学習器に適用されている場合には(1倍速再生時)、A/D変換部2のサンプリング周波数fsADは8KHZであり、D/A変換部8のサンプリング周波数fsDAは8KHZ

である。この場合には、ピッチ圧縮伸長手段23で、たとえば、2ピッチ周期が3ピッチ周期になるように、圧縮率3/2で音声信号が伸長される。つまり、音声区間が1.5倍に伸長される。したがって、この場合には、1倍速再生の通常再生時に対して、 $3/2 - 1 = 1/2$ だけ信号が伸長されることになり、この伸長分がリングメモリ7の蓄積量となる。

(2) 第2ケース(case 2)

入力信号が音声区間であり、かつリングメモリ7がオーバーフロー直前状態であると判別されたときには、第2ケースとなる。

【0061】この場合には、音声信号はマルチプレкса20を介して、入力信号削除部21に送られ、音声信号が削除される。具体的には、アップダウンカウンタ9のカウント値が、アンダーフロー検出用データTmin以下になるまで、すなわちリングメモリ7がアンダーフロー直前状態になるまで、リングメモリ7への書き込み動作が停止される。

【0062】リングメモリ7がアンダーフロー直前状態になると、200個以下の個数、例えば100個の消音信号(値"0"の信号)が消音挿入部22から出力され、この消音信号がデマルチプレкса27を介してリングメモリ7に送られて書き込まれる。このように、消音信号をリングメモリ7へ書き込んでいるのは、音声削除によって音声信号の二重目にクリック音が発生するのを防止するためである。

(3) 第3ケース(case 3)

入力信号が無音区間であり、かつ無音区間の継続長が設定されたポーズ継続長Tdel未満であり、かつリングメモリ7がオーバーフロー直前状態ではないと判別されたときには、第3ケースとなる。

【0063】この場合は、上記第1ケースの場合と同じ処理が行なわれる。ただし、第3ケースに該当する場合には、再生速度倍率をnとすると、 $1/n$ の圧縮率で伸長圧縮処理を行ってもよい。つまり、第3ケースに該当する場合には、 $1/n$ 以上の圧縮率で伸長圧縮処理が行われる。

(4) 第4ケース(case 4)

入力信号が無音区間であり、かつ無音区間の継続長が設定されたポーズ継続長Tdel未満であり、かつリングメモリ7がオーバーフロー直前状態であると判別されたときには、第4ケースとなる。

【0064】この場合は、上記第2ケースの場合と同じ処理が行なわれる。

(5) 第5ケース(case 5)

入力信号が無音区間であり、かつ無音区間の継続長が設定されたポーズ継続長Tdel以上であり、かつリングメモリ7がアンダーフロー直前状態ではないと判別されたときには、第5ケースとなる。

【0065】この場合には、音声信号はマルチプレкса

20を介して、入力信号削除部25に送られ、音声信号が削除される。具体的には、リングメモリ7への書き込み動作が停止される。ただし、音声区間のスタート部分（無声区間）が欠落するのを防止したり、音声の削除によって≡ぎ目にクリック音が発生したりするのを防止するために、波形合成挿入部26によって波形合成挿入処理が行なわれる。

【0066】波形合成挿入部26による波形合成挿入処理について、図4(a)、(b)を用いて説明する。図4(a)による方法では、波形合成挿入部26は、第1メモリ31および第2メモリ32を備えている。入力信号削除部25による入力信号削除処理の開始時においては、削除開始点から、1フレーム長以下の所定長さ T_s 、例えば1フレーム分の入力信号が、第1メモリ31にアドレス順に順次記憶される。次に、第1メモリ31のアドレスが大きくなるにしたがって1から0に直線的に変化する関数 K_1 が、第1メモリ31の内容 A に乗算される。そして、その乗算結果 A' が、再度第1メモリ31に書き込まれる。

【0067】また、入力信号削除部25による入力信号削除区間の終了点直前の所定長さ T_s 分の入力信号が、第2メモリ32にアドレス順に順次記憶される。次に、第2メモリ32のアドレスが大きくなるほど、0から1に直線的に変化する関数 K_2 が、第2メモリ32の内容 B に乗算される。そして、その乗算結果 B' が、再度第2メモリ32に書き込まれる。この後、第1メモリ31の内容 A' と、第2メモリ32の内容 B' とが加え合わされて、所定長さ T_s のデータ $A' * B'$ が得られる。そして、得られた所定長さ T_s 分のデータ $A' * B'$ がデマルチプレクサ27を介して、リングメモリ7に送られ、リングメモリ7に書き込まれる。

【0068】図4(b)による方法では、削除開始点から、1フレーム長以下の所定長さ T_s 、例えば1フレーム分の入力信号が、第1メモリ31にアドレス順に順次記憶される。次に、後端に1から0に直線的に変化するスロープがついた関数 K_3 が、第1メモリ31の内容 A に乗算される。そして、その乗算結果 A' が、再度第1メモリ31に書き込まれる。

【0069】また、入力信号削除部25による入力信号削除区間の終了点直前の所定長さ T_s 分の入力信号が、第2メモリ32にアドレス順に順次記憶される。次に、前端に0から1に直線的に変化するスロープがついた関数 K_4 が、第2メモリ32の内容 B に乗算される。そして、その乗算結果 B' が、再度第2メモリ32に書き込まれる。この後、第1メモリ31の内容 A' と、第2メモリ32の内容 B' とが≡ぎ合わされて、 $2T_s$ 分のデータ $A' + B'$ が得られる。そして、得られた $2T_s$ 分のデータ $A' + B'$ がデマルチプレクサ27を介して、リングメモリ7に送られ、リングメモリ7に書き込まれる。図4(b)では、 T_s が、1フレーム分の長さ

である例を示したが、1フレームの半分の長さのデータを T_s としてもよい。

【0070】なお、入力信号削除部25による無音区間の音声信号の削除処理が繰り返し行なわれている場合に、リングメモリ7がアンダーフロー直前状態になることがある。この場合には、リングメモリ7がアンダーフロー直前状態になったときから、所定長さ T_s 分の入力信号が第2メモリ32に記憶される。そして、第1メモリ31に記憶されているデータと、第2メモリ32に記憶されているデータにもとづいて、上記と同様な波形合成挿入処理が行なわれる。

(6) 第6ケース(case 6)

入力信号が無音区間であり、かつ無音区間の継続長が設定されたポーズ継続長 T_{del} 以上であり、かつリングメモリ7がアンダーフロー直前状態であると判別されたときには、第6ケースとなる。

【0071】この場合は、入力信号は、マルチプレクサ20を介して間引き処理部24に送られる。間引き処理部24では、VTRの再生速度倍率を n として、圧縮率が $1/n$ となるように間引き処理が行なわれる。たとえば、2倍速再生時には入力信号に対して圧縮率 $1/2$ で間引きが行なわれ、3倍速再生時には入力信号に対して圧縮率 $1/3$ で間引きが行なわれる。1倍速再生時には、入力信号がそのまま出力される。

【0072】 $1/n$ 間引き処理部24による間引き処理としては、次のような方法が用いられる。ここでは、2倍速再生時を例にとって説明する。

【0073】上述したPICOLAまたはTDHSを用いた時間軸圧縮法を用い、入力信号のピッチを抽出し、ピッチデータ部分を圧縮率が $1/2$ となるように、間引く。

【0074】また、図5(a)～(c)に示すように、ピッチ抽出をすることなく、所定時間 T_s ごとに波形を間引くようにしてもよい。

【0075】図5(a)の方法では、波形A～Dのうち、波形Bおよび波形Dが間引かれ、波形A、Cからなる信号が得られる。

【0076】図5(b)の方法では、波形A～Dのうち、波形Bと波形Dが間引かれている。また、波形Aには、前端に0から1に上昇するスロープ（関数 K_4 ）が、後端に1から0に下降するスロープ（関数 K_3 ）がついた関数が乗算されて、波形 A' が作成される。また、波形Cには、前端に0から1に上昇するスロープ（関数 K_4 ）が、後端に1から0に下降するスロープ（関数 K_3 ）がついた関数が乗算されて、波形 C' が作成される。このようにして、4つの波形A～Dからなる信号は、2つの波形 A' および C' からなる信号に圧縮される。

【0077】図5(c)の方法では、波形Aに対しては、1から0へ直線的に向かう重み（重み関数 K_1 ）が

つけられて、波形A'が作成される。波形Bに対しては0から1に向かう重み(関数K2)がつけられて、波形B'が作成される。そして、それらの波形A'およびB'が加え合わされ、長さTsの波形A' * B'が作成される。

【0078】同様に、波形Cに対しては、1から0へ直線的に向かう重み(関数K1)がつけられて、波形C'が作成される。波形Dに対しては0から1に向かう重み(関数K2)がつけられて、波形D'が作成される。そして、それらの波形C'およびD'が加え合わされ、長さTsの波形C' * D'が作成される。このようにして、4つの波形A~Dからなる信号は、2つの波形A' * B'およびC' * D'からなる信号に圧縮される。

【0079】上述のように、第6ケースに該当する場合には、VTRの再生倍率をnとして、圧縮率1/nで間引き処理が行われているが、次のようにして圧縮率を制御するようにしてもよい。

【0080】圧縮率1/nで間引き処理が行われている場合、D/A変換器8のサンプリング周波数fsDAとA/D変換器2のサンプリング周波数fsADとの比fsDA/fsADが、圧縮率1/nと等しい場合には、リングメモリ7の蓄積量は、変化しない。しかしながら、圧縮率1/nの演算精度、サンプリング周波数fsADとfsDAのクロック精度によっては、fsDA/fsADが圧縮率1/nと等しくなることが起こりうる。

【0081】fsDA/fsADが圧縮率1/nより大きくなったとき($fsDA/fsAD > 1/n$)には、 $fsDA/fsAD = 1/a$ ($a > 0$)として、 $\{(1/a) - (1/n)\}$ だけ、圧縮率が小さくなり、間引きの度合いが大きくなり、リングメモリ7の蓄積量が減少していき、リングメモリ7の蓄積量がアンダーフローするおそれがある。

【0082】一方、fsDA/fsADが圧縮率1/nより小さくなったとき($fsDA/fsAD < 1/n$)には、 $fsDA/fsAD = 1/a$ ($a > 0$)として、 $\{(1/n) - (1/a)\}$ だけ、圧縮率が大きくなり、間引きの度合いが小さくなり、リングメモリ7の蓄積量が増加していく。

【0083】したがって、間引き処理を行う場合には、リングメモリ7の蓄積量を確認して、次のように圧縮率を制御する。 $fsDA/fsAD = 1/a$ ($a > 0$)として、 $(1/n) - \alpha < 1/a < (1/n) + \alpha$ の条件を満たす α を選定する。ただし、 α は、0以上で1以下の値であり、例えば0.001~0.1の範囲の値である。

【0084】fsDA/fsADが圧縮率1/nより大きくなったとき、すなわち、リングメモリ7の蓄積量が減少していく場合には、圧縮率を1/nから $\{(1/n) + \alpha\}$ にする。つまり、圧縮率を大きくし、リング

メモリ7の蓄積量を増加させるようにする。

【0085】fsDA/fsADが圧縮率1/nより小さくなったとき、すなわち、リングメモリ7の蓄積量が増加していく場合には、圧縮率を1/nから $\{(1/n) - \alpha\}$ にする。つまり、圧縮率を小さくし、リングメモリ7の蓄積量を減少させるようにする。

【0086】上記では、リングメモリ7の蓄積量に基づいて、圧縮率を変化させているが、間引き処理が行われる場合に、フレーム毎に圧縮率を $\{(1/n) - \alpha\}$ または $\{(1/n) + \alpha\}$ に、交互に変化させるようにしてもよい。

【0087】図6および図7は、話速変換部6による処理手順を示している。

【0088】以下、VTRの2倍速再生時の場合の話速変換部6による処理について、説明する。

(1) 再生開始時の処理

再生が開始されて、パワー計算部11によって最初のフレームの平均パワー値Pが算出されると(ステップ1)、算出された平均パワー値Pがしきい値Th以上か否かが比較部12の出力に基づいて判別される(ステップ2)。

【0089】入力音声信号が無音区間から開始した場合、最初のフレームにおいては、平均パワー値Pはしきい値Thより小さくなり、ステップ1に進む。そして、無音区間の継続長(無音区間が継続するフレーム数)が算出され、算出された継続長がポーズ継続長メモリ17に設定されているポーズ継続長Tdel以上か否かが判別される(ステップ12)。このポーズ継続長Tdelは、たとえば、フレーム数にして4フレーム分の長さに設定されている。

【0090】最初のフレームに対する処理においては、無音区間の継続長がポーズ継続長Tdel未満であるので、リングメモリ蓄積量状態判別部16の出力に基づいて、リングメモリ7がアンダーフロー直前状態か否かが判別される(ステップ13、14)。

【0091】最初のフレームに対する処理においては、リングメモリ7は、アンダーフロー直前状態になっているので、フレームデータが間引き処理部24によって圧縮率1/2で間引かれ(ステップ28)、間引き処理後の圧縮データがリングメモリ7に書き込まれる。この後、ステップ1に戻る。

(2) 第1ケースとなる処理の説明

ステップ2で、平均パワー値Pがしきい値Th以上であると判別されたときには、今回のフレームが音声区間であると判断され、ステップ3に進む。ステップ3では、前フレームが削除区間であったか否かが、第1フラグF1の状態に基づいて判別される。前フレームが削除区間でない場合には、リングメモリ蓄積量状態判別部16の出力に基づいて、リングメモリ7がオーバーフロー直前状態か否かが判別される(ステップ6、7)。前フレー

ムが削除区間である場合には、ステップ4および5の処理が行なわれた後、リングメモリ7がオーバーフロー直前状態か否かが判別される(ステップ6、7)。ステップ4および5の処理については、後述する。

【0092】ステップ7において、オーバーフロー直前状態ではないと判別された場合には、第1ケースとなり、ピッチ圧縮伸長手段23によって、今回のフレームデータが2/3の圧縮率で時間軸圧縮される(ステップ8)。圧縮データは、リングメモリ7に送られて書き込まれる。この後、ステップ1に戻る。

(2) 第2ケースとなる処理の説明

ステップ2で、平均パワー値Pがしきい値Th以上であると判別されたときには、今回送られてきたフレームは音声区間であると判断され、ステップ3に進む。ステップ3では、前フレームが削除区間であったか否かが、第1フラグF1の状態に基づいて判別される。前フレームが削除区間でない場合には、リングメモリ蓄積量状態判別部16の出力に基づいて、リングメモリ7がオーバーフロー直前状態か否かが判別される(ステップ6、7)。前フレームが削除区間である場合には、ステップ4および5の処理が行なわれた後、リングメモリ7がオーバーフロー直前状態か否かが判別される(ステップ6、7)。ステップ4および5の処理については、後述する。

【0093】ステップ7において、オーバーフロー直前状態であると判別された場合には、第2ケースとなり、リングメモリ蓄積量状態判別部16からアンダーフロー検出信号が出力されるまで、入力信号削除部21によって入力信号が削除される(ステップ9)。つまり、リングメモリ7がアンダーフロー直前状態になるまで、リングメモリ7への書き込みが停止される。

【0094】そして、リングメモリ7がアンダーフロー直前状態になると、消音挿入部22によって、200個以下の所定数の消音信号“0”がリングメモリ7に書き込まれる(ステップ10)。そして、ステップ1に戻る。

【0095】上記ステップ10の処理の代わりに、図9(a)または図9(b)に示すような処理を行なってもよい。図9(a)に示す方法について説明すると、ステップ7でオーバーフロー直前状態と判別されたときから、たとえば、200個の入力信号に対する波形Aに対しては、1から0へ直線的に向かう重み(重み関数K1)をつけて波形A'を得る。また、アンダーフロー直前から200個前までの200個の入力信号に対する波形Bに対しては0から1に向かう重み(重み関数K2)をつけて、波形B'を得る。

【0096】そして、得られた2つの波形A'およびB'を加え合わせて、200個分の長さの波形A' * B'を作成する。そして、この波形A' * B'に対する200個の信号をリングメモリ7に書き込む。なお、ア

ンダーフロー直前から200個前の時点の検出は、アップダウンカウンタ9のカウント値に基づいて行なわれる。これにより、音声削除区間の前後の音声信号の二重目にクリック音が発生するのを、効果的に防止できる。

【0097】図9(b)に示す方法について説明すると、ステップ7でオーバーフロー直前状態と判別されたときから、たとえば、100個の入力信号に対する波形Aに対しては、1から0へ直線的に向かう重み(重み関数K1)をつけて波形A'を得る。また、アンダーフロー直前から100個前までの100個の入力信号に対する波形Bに対しては0から1に向かう重み(重み関数K2)をつけて、波形B'を得る。そして、得られた2つの波形A'およびB'を二重合わせた200個分の信号をリングメモリ7に書き込む。

【0098】上記ステップ9では、オーバーフロー直前状態であると判別された場合には、リングメモリ蓄積量状態判別部16からアンダーフロー検出信号が出力されるまで、入力信号削除部21によって入力信号が削除されているが、リングメモリ7に蓄積されているデータを、リングメモリ7がアンダーフロー直前状態になるように、削除するようにしてもよい。

【0099】具体的には、リングメモリ7の書込開始アドレスを、図18(a)に示すオーバーフロー直前状態の時のアドレス(C地点)から、図18(b)に示すようにリングメモリ7がアンダーフロー直前状態となるアドレス(A地点)までジャンプさせる。したがって、ステップ9の処理では、A地点からC地点までのアドレスに蓄積されていたデータが削除されることになる。この後、図18(c)に示すように、ステップ10によって消音信号が書き込まれた後、入力データが書き込まれていく。

【0100】ステップ9において、上記のように、リングメモリ7に蓄積されているデータをリングメモリ7がアンダーフロー直前状態になるように削除した場合、ステップ10で消音信号をリングメモリ7に書き込む代わりに図19(a)または図19(b)のような処理を行ってもよい。

【0101】今、リングメモリ7の書込開始アドレスが、図18(a)に示すオーバーフロー直前状態の時のアドレス(C地点)から、図18(b)に示すようにリングメモリ7がアンダーフロー直前状態となるアドレス(A地点)までジャンプしたとする。このA地点から所定数、例えば200先のアドレス(図19(a)のB地点)までに蓄積されているデータSに対しては、図19(a)に示すように、1から0へ直線的に向かう重み(重み関数K1)をつけて波形S'を得る。また、それ以後にリングメモリ7に書き込まれる200個分の入力データ(波形T)に対しては、図19(a)に示すように、0から1に向かう重み(重み関数K2)をつけて、波形T'を得る。

【0102】そして、得られた2つの波形S'およびT'を加え合わせて、200個分の長さの波形S' * T'を作成する。そして、この波形S' * T'に対する200個の信号をA地点からリングメモリ7に書き込む。これにより、蓄積データ削除区間の前後の音声信号の二重目にクリック音が発生するのを、効果的に防止できる。

【0103】図19(b)に示す方法について説明すると、図18(b)のA地点から所定数、例えば100個先のアドレス(図19(b)のB地点)までに蓄積されているデータSに対しては、1から0へ直線的に向かう重み(重み関数K1)をつけて波形S'を得る。また、それ以後にリングメモリ7に書き込まれる100個分の入力データ(波形T)に対しては、0から1に向かう重み(重み関数K2)をつけて、波形T'を得る。そして、得られた2つの波形S'およびT'を二重合わせた200個分の信号をA地点からリングメモリ7に書き込む。

(3) 第3ケースとなる処理の説明

ステップ2で平均パワー値Pがしきい値Thより小さいと判別されたときには、今回までの無音区間の継続長が算出され(ステップ11)、算出された継続長がポーズ継続長メモリ17に設定されているポーズ継続長Tde1以上か否かが判別される(ステップ12)。そして、無音区間の継続長がポーズ継続長Tde1未満であると判別された場合には、リングメモリ蓄積量状態判別部16の出力に基づいて、アンダーフロー直前状態か否かが判別される(ステップ13、14)。

【0104】リングメモリ7がアンダーフロー直前状態になっていないときには、リングメモリ蓄積量状態判別部16の出力に基づいて、オーバーフロー直前状態か否かが判別される(ステップ6、7)。オーバーフロー直前状態でない場合には、第3ケースとなり、ピッチ圧縮伸長手段23によって、今回のフレームデータが2/3の圧縮率で時間軸圧縮される(ステップ8)。圧縮データは、リングメモリ7に送られて書き込まれる。この後、ステップ1に戻る。

(4) 第4ケースとなる処理の説明

ステップ2で平均パワー値Pがしきい値Thより小さいと判別されたときには、今回までの無音区間の継続長が算出され(ステップ11)、算出された継続長がポーズ継続長メモリ17に設定されているポーズ継続長Tde1以上か否かが判別される(ステップ12)。そして、無音区間の継続長がポーズ継続長Tde1未満であると判別された場合には、リングメモリ蓄積量状態判別部16の出力に基づいて、アンダーフロー直前状態か否かが判別される(ステップ13、14)。

【0105】リングメモリ7がアンダーフロー直前状態になっていないときには、リングメモリ蓄積量状態判別部16の出力に基づいて、オーバーフロー直前状態か否

かが判別される(ステップ6、7)。オーバーフロー直前状態である場合には、第4ケースとなり、リングメモリ蓄積量状態判別部16からアンダーフロー検出信号が出力されるまで、入力信号削除部21によって入力信号が削除される(ステップ9)。つまり、リングメモリ7がアンダーフロー直前状態になるまで、リングメモリ7への書き込みが中断される。

【0106】そして、リングメモリ7がアンダーフロー直前状態になると、消音挿入部22によって、200個以下の所定数の消音信号"0"がリングメモリ7に書き込まれる(ステップ10)。そして、ステップ1に戻る。

(5) 第5ケースとなる処理の説明

ステップ2で平均パワー値Pがしきい値Thより小さいと判別されたときには、今回までの無音区間の継続長が算出され(ステップ11)、算出された継続長がポーズ継続長メモリ17に設定されているポーズ継続長Tde1以上か否かが判別される(ステップ12)。そして、無音区間の継続長がポーズ継続長Tde1以上であると判別された場合には、リングメモリ蓄積量状態判別部16の出力に基づいて、アンダーフロー直前状態か否かが判別される(ステップ15、16)。

【0107】リングメモリ7がアンダーフロー直前状態でないときには、第5ケースとなり、今回のフレームが入力信号削除部25による削除区間であることを示す第1フラグF1がセットされる(ステップ17)。この第1フラグF1は、電源投入時の初期設定において、リセット(F1=0)されている。そして、今回のフレームが入力信号削除部25による削除区間の最初のフレームであるか否かを示す第2フラグF2がリセットされているか否かが判別される(ステップ18)。

【0108】この第2フラグF2は、電源投入時の初期設定において、リセット(F2=0)されている。そして、入力信号削除部25による削除区間の最初のフレームに対する処理が終了したときにセット(F2=1)にされる。そして、入力信号削除部25による一連の削除区間に対する処理が終了したときにリセット(F2=0)される。

【0109】したがって、今回のフレームが入力信号削除部25による削除区間の最初のフレームであるときには、第2フラグF2は、リセット(F2=0)されている。第2フラグF2がリセットされているときには、波形合成挿入部26によって第1メモリ31に今回のフレームデータが記憶される(ステップ19)。また、入力信号削除部25によって今回のフレームデータのリングメモリ7への書き込みが停止される(ステップ20)。つまり、今回のフレームデータが削除される。そして、第2フラグF2がセット(F2=1)された後(ステップ21)、ステップ1に戻る。

【0110】さらに、無音区間が続いている場合には、

ステップ2、11、12、15を通してステップ16に移り、リングメモリ蓄積量状態判別部16の出力に基づいて、リングメモリ7がアンダーフロー直前状態か否かが判別される。

【0111】リングメモリ7がアンダーフロー直前状態でないときには、今回のフレームが入力信号削除部25による削除区間であることを示す第1フラグF1がセットされる(ステップ17)。そして、今回のフレームが入力信号削除部25による削除区間の最初のフレームであるか否かを示す第2フラグF2がリセットされているか否かが判別される(ステップ18)。

【0112】この場合には、第2フラグF2はセット($F2=1$)されているので、今回のフレームが入力信号削除部25による削除区間の最初のフレームでないと判断される。この場合には、波形合成挿入部26によって第2メモリ32に今回のフレームデータが記憶される(ステップ22)。また、入力信号削除部25によって今回のフレームデータのリングメモリ7への書き込みが停止される(ステップ23)。そして、ステップ1に戻る。

【0113】そして、さらに、無音区間が続きかつリングメモリ7がアンダーフロー直前状態となっていないときには、ステップ2、11、12、15、16、17、18、22および23の処理が繰り返される。つまり、第2メモリ32のフレームデータが更新されるとともに、フレームデータのリングメモリ7への書き込みが停止される。

【0114】この後、音声区間のフレームデータが入力されたときには、ステップ2において、平均パワー値Pがしきい値 T_h 以上となるので、前フレームが入力信号削除部25による削除区間であったか否かが、第1フラグF1状態に基づいて判別される(ステップ3)。この場合には、第1フラグF1がセット($F1=1$)されているので、前フレームが入力信号削除部25による削除区間であったと判別され、ステップ4に移る。ステップ4では、入力信号削除部25による削除処理が停止せしめられるとともに、波形合成挿入部26による波形合成挿入処理が行なわれる。

【0115】すなわち、図4(a)を用いて既に説明したように、第1メモリ31の内容に1から0に直線的に変化する関数が乗算され、第2メモリ32の内容に0から1に直線的に変化する関数が乗算され、これらの両乗算結果が加え合わされる。この加算結果(図4(a)の $A' \cdot B'$ に相当する。)が、デマルチプレクサ27を介して、リングメモリ7に送られ、リングメモリ7に書き込まれる。

【0116】この後、第1フラグF1および第2フラグF2がリセット($F1=F2=0$)される(ステップ5)。そして、ステップ6に進む。

【0117】ところで、連続している無音区間に対し

て、上記のような入力信号削除部25による削除処理が繰り返行なわれている場合において、リングメモリ7がアンダーフロー直前状態になることがある。この場合には、上記ステップ16でYESとなり、ステップ24に移る。ステップ24では、前フレームが入力信号削除部25による削除区間であったか否かが、第1フラグF1の状態に基づいて判別される。

【0118】この場合には、第1フラグF1がセット($F1=1$)されているので、ステップ25に進み、第2メモリ32に今回のフレームデータが記憶される。そして、入力信号削除部25による削除処理が停止せしめられるとともに、波形合成挿入部26による波形合成挿入処理が行なわれる(ステップ26)。そして、第1フラグF1および第2フラグF2がリセット($F1=F2=0$)された後(ステップ27)、ステップ1に進む。

【0119】上記ステップ26における波形合成挿入部26による波形合成挿入処理には、上記ステップ4で説明した波形合成挿入処理とほぼ同様であるが、第2メモリ32に記憶されているフレームデータが、リングメモリ7がアンダーフロー直前状態になった後のフレームデータである点が、上記ステップ4で説明した処理の場合と異なっている。

【0120】なお、上記ステップ25の処理を省略し、ステップ24でYESとなった場合に、第2メモリ32に今回のフレームデータを記憶させることなく、ステップ26に移るようにしてもよい。この場合には、ステップ26で行なわれる波形合成挿入処理においては、上記ステップ4で説明した波形合成挿入処理と同様に、第2メモリ32に記憶されているアンダーフロー直前状態より前のフレームデータ(前回のフレームデータ)が用いられる。

【0121】また、上記ステップ22の処理を省略するとともに上記ステップ3と上記ステップ4との間に、フレームデータを第2メモリ32に記憶させるステップを追加するようにしてもよい。この場合には、ステップ4においては、上記ステップ19において第1メモリ31に記録された内容と、上記ステップ3と上記ステップ4との間に追加されたステップにおいて第2メモリ32に記録された内容とに基づいて、波形合成挿入処理が行われる。

(6) 第6ケースとなる処理の説明

ステップ2で平均パワー値Pがしきい値 T_h より小さいと判別されたときには、今回までの無音区間の継続長が算出され(ステップ11)、算出された継続長がポーズ継続長メモリ17に設定されているポーズ継続長 T_{de1} 以上か否かが判別される(ステップ12)。そして、無音区間の継続長がポーズ継続長 T_{de1} 以上であると判別された場合には、リングメモリ蓄積量状態判別部16の出力に基づいて、アンダーフロー直前状態か否かが判別される(ステップ15、16)。

【0122】リングメモリ7がアンダーフロー直前状態であるときには、前フレームが入力信号削除部25による削除区間であったか否かが、第1フラグF1の状態に基づいて判別される(ステップ24)。第1フラグF1がリセットされている場合($F1=0$)、すなわち、前フレームが入力信号削除部25による削除区間でなかった場合には、第6ケースとなり、ステップ28に移る。ステップ28では、間引き処理部24によって、今回のフレームデータが圧縮率 $1/2$ で間引き処理が行なわれる。そして、間引き処理されたデータは、リングメモリ7に送られて書き込まれる。この後、ステップ1に戻る。

【0123】つまり、無音区間の継続長がポーズ継続長 T_{del} 以上であっても、リングメモリ7がアンダーフロー直前状態であり、かつ前フレームが入力信号削除部25による削除区間でない場合には、フレームデータは削除されず、圧縮率 $1/2$ で間引き処理が行なわれた後、リングメモリ7に書き込まれる。

【0124】図7においては、ステップ12において、無音区間の継続長が設定されたポーズ継続長 T_{del} より長いかが判別されているが、図8のステップ12Aに示すように、無音区間の継続長 T が設定された第1基準長 $T1$ 未満か($T < T1$)、無音区間の継続長 T が設定された第1基準長 $T1$ 以上で設定された第2基準長 $T2$ (ただし $T1 < T2$)未満か($T1 \leq T < T2$)、または無音区間の継続長 T が設定された第2基準長 $T2$ 以上か($T \geq T2$)を、判別するようにしてもよい。第1基準長としては、たとえば、4フレーム分の長さが、第2基準長としてはたとえば40フレーム分の長さが設定される。

【0125】そして、図8に示すように、各判別結果に応じて、次のようなステップに進むようにしてもよい。すなわち、無音区間の継続長 T が設定された第1基準長 $T1$ 未満($T < T1$)である場合には、ステップ13に進む。無音区間の継続長 T が設定された第1基準長 $T1$ 以上で設定された第2基準長 $T2$ ($T1 < T2$)未満($T1 \leq T < T2$)であるときには、ステップ28に進んで $1/n$ 間引き処理による間引きを行なう。無音区間の継続長 T が設定された第2基準長 $T2$ 以上($T \geq T2$)であるときには、ステップ15に進む。

【0126】図10は、2倍速再生時の入力信号と出力信号との関係を示し、特に無音区間の入力信号が削除される様子を示している。図11および図12は、リングメモリ7へのデータ書き込み開始点、リングメモリ7からのデータ読み出し開始点ならびに図10の各点A~Hにおけるリングメモリ7の状態を示している。

【0127】図10では、2倍速再生開始時においては、入力信号は無音区間となっており、かつリングメモリ7は空状態であるので(図11(a)参照)、フレームデータが間引き処理部24によって圧縮率 $1/2$ で間

引かれた後、リングメモリ7に書き込まれていく。

【0128】そして、リングメモリ7の蓄積量 T_m がアンダーフロー検出用データ T_{min} に達すると、リングメモリ7からのデータの読み出しが開始される(図11(b)参照)。

【0129】そして、入力信号の音声区間aに対するフレームデータが送られてくると(A点)、ピッチ圧縮伸長手段23によって、フレームデータが圧縮率 $2/3$ で圧縮される。入力信号と出力信号との長さが一致する圧縮率 $1/2$ の圧縮を基準とすると、フレームデータが伸長される。この意味で、図10には、伸長処理と記載されている。そして、この圧縮データがリングメモリ7に書き込まれる。A点においては、図11(c)に示すように、蓄積量 T_mA は、 T_{min} のままである。

【0130】入力信号の音声区間aに対する出力信号a1は、A点での蓄積量 T_mA 分だけ遅れて読み出されていく。そして、入力信号の音声区間aが入力され終わった時点(B点)では、図11(d)に示すように、今回の圧縮区間の開始点であるA点での蓄積量 T_{min} と、A点からB点までの音声区間aの圧縮データの、圧縮率 $1/2$ の圧縮に対する伸長分 S_{tB} との和がリングメモリ7の蓄積量 $T_mB (=S_{tB} + T_{min})$ となる。したがって、入力信号の音声区間aに対する出力信号a1は、B点から $T_mB (=S_{tB} + T_{min})$ 分が経過した点で出力され終わる。

【0131】入力信号の音声区間aに続くポーズ継続長 T_{del} 未満の無音区間のフレームデータも、ピッチ圧縮伸長手段23によって圧縮率 $2/3$ で圧縮される。この無音区間に続いて音声区間bが入力されると、この音声区間bのフレームデータもピッチ圧縮伸長手段23によって圧縮率 $2/3$ で圧縮される。

【0132】そして、入力信号の音声区間bが入力され終わった時点(C点)では、図11(e)に示すように、今回の圧縮区間の開始点であるA点での蓄積量 T_{min} と、A点からC点までの入力信号に対応する圧縮データの、 $1/2$ 圧縮に対する伸長分 S_{tC} との和がリングメモリ7の蓄積量 $T_mC (=S_{tC} + T_{min})$ となる。したがって、入力信号の音声区間bに対する出力信号b1は、C点から $T_mC (=S_{tC} + T_{min})$ 分が経過した点で出力され終わる。

【0133】入力信号の音声区間bに続いて、ポーズ継続長 T_{del} 以上の長さの無音区間の信号が送られてきたときには、ポーズ継続長 T_{del} に達するまで(D点)はフレームデータが、ピッチ圧縮伸長手段23によって圧縮率 $2/3$ で圧縮される。

【0134】D点では、図11(f)に示すように、今回の圧縮区間の開始点であるA点での蓄積量 T_{min} と、A点からD点までの入力信号に対応する圧縮データの、 $1/2$ 圧縮に対する伸長分 S_{tD} との和がリングメモリ7の蓄積量 $T_mD (=S_{tD} + T_{min})$ となる。

したがって、入力信号の音声区間bとD点との間の無音区間に対する出力信号は、D点から $TmD (=StD + Tmin)$ 分が経過した点で出力され終わる。

【0135】ポーズ継続長 $Tdel$ 以降の無音区間のフレームデータは、リングメモリ7の蓄積量がアンダーフロー検出用データ $Tmin$ 以下になるまで、入力信号削除部25によって削除される。このポーズ削除部分の長さ Std は、今回の圧縮区間の開始点であるA点からD点までの入力信号に対応する圧縮データの、 $1/2$ 圧縮に対する伸長分 StD と等しくなる。入力信号削除部25によって削除処理が行なわれた後においては、波形合成挿入部22によってクリック音防止のための合成波形が挿入されるが、図10には挿入された合成波形部分を省略してある。

【0136】入力信号が削除された区間の最終点(E点)においては、図12(g)に示すように、リングメモリ7の蓄積量 TmE は、アンダーフロー検出用データ $Tmin$ 以下となる。ここでは、蓄積量 TmE が、アンダーフロー検出用データ $Tmin$ に等しくなった例を示している。

【0137】E点からの無音区間に対するフレームデータは、間引き処理部24によって、圧縮率 $1/2$ で間引かれた後、フレームメモリ7に書き込まれる。そして、音声区間cの信号が入力されると(F点)、この音声区間cのフレームデータがピッチ圧縮伸長手段23によって、圧縮率 $2/3$ で圧縮される。つまり、新たな圧縮区間が開始される。そして、圧縮データがリングメモリ7に書き込まれる。

【0138】F点では、図12(h)に示すように、リングメモリ7の蓄積量 TmF は、E点のときと同じ $Tmin$ となっている。

【0139】入力信号の音声区間cに対する出力信号 $c1$ は、F点での蓄積量 $Tmin$ 分だけ遅れて出力されていく。入力信号の音声区間cに続くポーズ継続長 $Tdel$ 未満の無音区間(音声区間cからG点までの無音区間)のフレームデータも、ピッチ圧縮伸長手段23によって圧縮率 $2/3$ で圧縮される。

【0140】G点では、図12(i)に示すように、今回の圧縮区間の開始点であるF点での蓄積量 $Tmin$ と、F点からG点までの入力信号に対応する圧縮データの、 $1/2$ 圧縮に対する伸長分 StG との和がリングメモリ7の蓄積量 $TmG (=StG + Tmin)$ となる。したがって、入力信号の音声区間cからG点までの無音区間に対する出力信号は、G点から $TmG (=StG + Tmin)$ 分が経過した点で出力され終わる。

【0141】ポーズ継続長 $Tdel$ 以降の無音区間のフレームデータは、リングメモリ7の蓄積量がアンダーフロー検出用データ $Tmin$ になるまで、入力信号削除部25によって削除される。このポーズ削除部分の長さ Std は、今回の圧縮区間の開始点であるF点からG点ま

での入力信号に対応する圧縮データの、 $1/2$ 圧縮に対する伸長分 StG と等しくなる。

【0142】入力信号が削除された区間の最終点(H点)においては、図12(j)に示すように、リングメモリ7の蓄積量 TmH は、アンダーフロー検出用データ $Tmin$ 以下となる。ここでは、蓄積量 TmH が、アンダーフロー検出用データ $Tmin$ に等しくなった例を示している。

【0143】H点からの無音区間に対するフレームデータは、間引き処理部24によって、圧縮率 $1/2$ で間引かれた後、フレームメモリ7に書き込まれる。そして、音声区間dの信号が入力されると、この音声区間dのフレームデータがピッチ圧縮伸長手段23によって、圧縮率 $2/3$ で圧縮される。そして、伸長されたデータがリングメモリ7に書き込まれる。

【0144】図13は、2倍速再生時の入力信号と出力信号との関係を示し、特にオーバーフロー直前状態となったときに、入力信号が削除される様子を示している。図14は、図13の各点S~Uにおけるリングメモリ7の状態を示している。

【0145】ある時点からT点までの、音声区間a、b、c等と無音区間とを含む一連の入力信号に対するフレームデータが、ピッチ圧縮伸長手段23によって圧縮率 $2/3$ で圧縮され(圧縮率 $1/2$ の圧縮に対しては伸長され)ているとする。この場合には、リングメモリ7に伸長分が蓄積されていく。

【0146】音声区間bの入力開始点(S点)においては、図14(a)に示すように、当該1連の入力信号の圧縮処理の開始点での蓄積量 $Tmin$ と、上記圧縮処理の開始点からS点までの入力信号に対応する圧縮データの、 $1/2$ 圧縮に対する伸長分 StS との和がリングメモリ7の蓄積量 $TmS (=StS + Tmin)$ となる。したがって、音声区間bに対する出力信号 $b1$ は、S点から $TmS (=StS + Tmin)$ 分が経過した点で出力され始められる。

【0147】音声区間cの入力信号に対応する圧縮データがリングメモリ7に書き込まれた時点(T点)において、リングメモリ7がオーバーフロー直前状態になったとする。すなわち、T点において、リングメモリ7の蓄積量がオーバーフロー検出用データ $Tmax$ 以上になったとする。

【0148】T点においては、図14(b)に示すように、当該1連の入力信号に対する圧縮処理の開始点での蓄積量 $Tmin$ と、上記圧縮処理開始点からT点までの入力信号に対応する圧縮データの、 $1/2$ 圧縮に対する伸長分 StT との和がリングメモリ7の蓄積量 $TmT (=StT + Tmin)$ となる。言い換えれば、リングメモリ7の全ワード数をTOTALとし、オーバーフロー検出用データを $Tmax$ とし、TOTALと $Tmax$ との差を $Dmin$ とすると、T点での蓄積量 Tmt は、

T_{max} に等しいので、 $TOTAL-Dmin$ となる。

【0149】したがって、当該1連の入力信号に対する出力信号は、T点から蓄積量 $TmT (=StT + Tmin)$ 分遅れた時点で出力され終わる。

【0150】T点において、リングメモリ7がオーバーフロー直前状態になると、それ以後の入力信号に対しては、リングメモリ7がアンダーフロー直前状態になるまで、入力信号削除部21によって無条件に削除される。入力信号削除部21によって削除処理が行なわれた後においては、消音挿入部22によって消音が挿入されるが、図13には挿入された消音部分を省略してある。リングメモリ7がオーバーフロー直前状態になった後(T点)、フレームデータが削除されていき、図14(c)に示すようにU点でリングメモリ7がアンダーフロー直前状態(蓄積量 $TmU = Tmin$)になったとする。この場合には、T点からU点までの4つの無音区間および3つの音声区間d、e、fからなる入力信号が削除される。したがって、T点からU点までの入力信号は、出力信号としては現れない。

【0151】U点の後に音声区間gの信号が入力されると、この音声区間に対するフレームデータは、ビット圧縮伸長手段23によって圧縮率2/3で圧縮され(圧縮率1/2の圧縮に対しては伸長され)た後、リングメモリ7に書き込まれていく。音声区間gに対する出力信号gは、U点でのリングメモリ7の蓄積量 $Tmin$ 分だけ遅れて出力され始められる。

【0152】上記実施例では、入力信号の音声区間と無音区間とを、各フレームの平均パワー値Pに基づいて判別しているが、各フレームの平均振幅に基づいて判別するようにしてもよい。この場合には、図15に示すように、図2のパワー計算部11の代わりにフレーム単位で平均振幅値を計算する平均振幅計算部11Aが設けられ、しきい値メモリ13Aには、A/D変換部2の量子化ビット数が12bitのときには、たとえば、値26のしきい値が設定される。そして、平均振幅計算部11Aによって計算された平均振幅値と、しきい値メモリ13Aのしきい値とが、比較部12Aによって比較されることにより、音声区間か無音区間かが判別される。

【0153】つまり、平均振幅値がしきい値以上であれば音声区間と判別され、平均振幅値がしきい値未満であれば無音区間と判別される。フレーム単位の平均振幅値Wは、サンプリングされた1フレーム内の各音声信号の振幅を i_0, i_1, \dots, i_{N-1} (ただし、 $N=200$)とすると、次の数式3に基づいて算出される。

【0154】

【数3】

$$W = (1/N) \times \sum_{k=0}^{N-1} |i_k|$$

【0155】その他の処理については、図2の話速変換部6による処理と同じであるので、その説明を省略する。

【0156】なお、この場合においても、次のようにして、しきい値を変更するようにしてもよい。すなわち、図15に点線で示すように、平均振幅定常状態検出およびしきい値更新部14Aを設ける。平均振幅定常状態検出およびしきい値更新部14Aは、平均振幅計算部11Aからの平均振幅値Wが、所定フレーム数にわたって一定であったか否かを判別し、一定であったときには(定常状態)、そのときの平均振幅値Wの2倍の値をしきい値メモリ13Aに書き込み、しきい値を更新させる。ただし、更新されるしきい値の最大値は、所定値、たとえば28に制限される。

【0157】また、入力信号の音声区間と無音区間とを、次の数式4で示す各フレームの音声信号の振幅累積値 Wa と所与のしきい値とに基づいて判別するようにしてもよい。

【0158】

【数4】

$$k = N-1 \\ Wa = \sum_{k=0}^N |i_k| \\ k = 0$$

【0159】また、入力信号の音声区間と無音区間とを、各フレームの信号の周期性を検出し、検出した周期が予め定められた音声信号のビット周期範囲内であれば、音声区間であると判別し、検出した周期が予め定められた音声信号のビット周期範囲外であれば無音区間であると判別するようにしてもよい。

【0160】この場合には、図16に示すように、図2のパワー計算部11の代わりに、自己相関法に基づいて、フレームごとの周期性を検出するビット周期検出部11Bが設けられ、しきい値メモリ13Bには、音声信号のビット周期範囲が設定される。そして、ビット周期検出部11Bで検出された周期と、しきい値メモリ13Bに設定された音声信号のビット周期範囲とが、比較部12Bによって比較される。

【0161】設定される音声信号のビット周期範囲は、再生速度により異なり、n倍速再生のときには、たとえば、 $66 \times n$ (Hz) ~ $320 \times n$ (Hz) の範囲に設定される。したがって、2倍速再生時には、音声信号のビット周期範囲は、132 Hz ~ 640 Hz の範囲に設定される。その他の処理については、図2の話速変換部6による処理と同じであるので、その説明を省略する。

【0162】また、入力信号の音声区間と無音区間とを、各フレームの信号のパワースペクトルと、定常状態のパワースペクトルと比較することにより、判別するようにしてもよい。

【0163】この場合には、図20に示すように、図2

のパワー計算部11の代わりに、フレームごとに所定の1または複数の周波数帯域に対するパワースペクトルを算出するパワースペクトル算出部11Cが設けられる。また、上記所定の1または複数の周波数帯域に対する定常状態のパワースペクトルがパワースペクトル記憶部13Cに記憶されている。

【0164】パワースペクトル記憶部13Cの内容は、パワースペクトル算出部11Cによって算出されたパワースペクトルの変化状態に基づいて、パワースペクトル定常状態検出部14Bが定常状態であることを検出したときには、検出された定常状態でのパワースペクトルに更新される。

【0165】入力信号がパワースペクトル算出部11Cに送られてくると、フレームごとに所定の1または複数の周波数帯域に対するパワースペクトルが算出される。そして、算出されたパワースペクトルと、パワースペクトル記憶部13Cに記憶されている定常状態のパワースペクトルと比較部12Cによって比較される。

【0166】算出されたパワースペクトルが定常状態のパワースペクトルに対して、変動していれば、そのフレームは音声区間と判別される。逆に、算出されたパワースペクトルが定常状態のパワースペクトルに対して、変動していなければ、そのフレームは無音区間と判別される。

【0167】具体的には、パワースペクトル記憶部13Cには、上記所定の1または複数の周波数帯域に対する定常状態のパワースペクトルに基づいて、上記所定の1または複数の周波数帯域に対するしきい値が記憶される。そして、パワースペクトル記憶部13Cに記憶されているパワースペクトル算出部11Cによって算出された上記所定の1または複数の周波数帯域に対するパワースペクトルと、パワースペクトル記憶部13Cに記憶されている対応するしきい値と比較されることにより、入力信号が音声区間か無音区間かが判別される。

【0168】たとえば、定常状態のパワースペクトルが図21の(a)に示されているように、雑音のみのパワースペクトルであるとする。また、雑音が含まれていない音声のパワースペクトルが図21の(b)に示されているものとする。定常状態において、図21(a)のパワースペクトルで示される雑音が存在する場合に、図21(b)で示すパワースペクトルを持つ音声信号が入力すると、そのパワースペクトルは、図21(c)に示されるように、両者のパワースペクトルが合成されたものとなる。

【0169】したがって、たとえば、定常状態のパワー

スペクトルにおいてパワーが比較的小さい周波数帯域f aおよびf bに対するパワーは、音声区間のパワースペクトルにおいては大幅に増加する。つまり、定常状態のパワースペクトルにおいてパワーが比較的小さい1または複数の周波数帯域における定常状態のパワーと、入力信号のパワースペクトルの上記1または複数の周波数帯域におけるパワーとを比較することにより、入力信号が音声区間か無音区間かを判別することができる。

【0170】なお、定常状態の雑音が高い周波数帯域の雑音であると判明している場合には、雑音の影響の少ない低い周波数帯域(例えば、4KHz以下の周波数帯域)に対するパワースペクトルを算出し、算出されたパワースペクトルが所定のしきい値以上か否かによって、入力信号が音声区間か無音区間かを判別することもできる。

【0171】また、各フレームのパワー平均値Pと、しきい値Thとを比較することにより、音声区間と無音区間とを判別する場合において、リングメモリ7の蓄積量に基づいて、しきい値Thを変化させるようにしてもよい。すなわち、リングメモリ7の蓄積量が少なくなるほど、言い換えれば、リングメモリ7の空領域が多くなるほど、音声区間の欠落部が少なくなるようにしきい値Thは小さくされる。これにより、出力音声が自然により近くなる。

【0172】つまり、図22に示すように、しきい値調整手段51を設ける。しきい値調整手段51は、リングメモリ蓄積量状態判別部16からリングメモリ7の蓄積量を得る。そして、得られたリングメモリ7の蓄積量を、D/A変換部8のサンプリング周波数で除することにより、蓄積時間Tmを算出する。そして、算出された蓄積時間Tmに基づいて、しきい値Thを決定し、しきい値メモリ13の内容を更新する。

【0173】より具体的に説明すると、リングメモリ蓄積量状態判別部16から得られたリングメモリ7の蓄積量がD/A変換部8のサンプリング周波数である8000で除されることにより、蓄積時間Tmが求められる。そして、予め作成された蓄積時間Tmに対するしきい値Thのデータに基づいて、蓄積時間Tmに対するしきい値Thが求められる。

【0174】次の表は、A/D変換部2の量子化ビット数が12bitである場合における蓄積時間Tmに対するしきい値Thのデータの一例を示している。

【0175】

【表1】

T _m	0.25～ 0.5 秒	0.5 ～ 0.75秒	0.75～ 1.0 秒	1.0 ～ 1.25秒	1.25～ 1.5 秒	1.5 ～ 1.75秒	1.75～ 2.5 秒	2.5 秒 以上
T _h	2 ⁷	2 ⁸	2 ⁹	2 ¹⁰	2 ¹¹	2 ¹²	2 ¹³	2 ¹⁴

【0176】また、各フレームのパワー累積値P_aとしきい値とを比較することにより、音声区間と無音区間とを判別する場合、各フレームの平均振幅値Wとしきい値とを比較することにより、音声区間と無音区間とを判別する場合、各フレームの振幅累積値W_aとしきい値とを比較することにより、各フレームのパワーストックとしきい値とを比較することにより、音声区間と無音区間とを判別する場合にも、上記と同様に、リングメモリ7の蓄積量に基づいて、しきい値を変化させるようにしてもよい。

【0177】また、リングメモリ7の蓄積量に基づいて、無音区間の削除開始点を決定するためのポーズ継続長T_{de1}を変化させるようにしてもよい。すなわち、リングメモリ7の蓄積量が少なくなるほど、言い換えれば、リングメモリ7の空領域が多くなるほど、無音区間の削除部が少なくなるように、ポーズ継続長T_{de1}が長くされる。これにより、出力音声が自然により近くなる。

【0178】つまり、図22に示すように、ポーズ継続長調整手段52を設ける。ポーズ継続長調整手段52

は、リングメモリ蓄積量状態判別部16からリングメモリ7の蓄積量を得る。そして、得られたリングメモリ7の蓄積量を、D/A変換部8のサンプリング周波数で除することにより、蓄積時間T_mを算出する。そして、算出された蓄積時間T_mに基づいて、ポーズ継続長T_{de1}を決定し、ポーズ継続長設定メモリ17の内容を更新する。

【0179】より具体的に説明すると、リングメモリ蓄積量状態判別部16から得られたリングメモリ7の蓄積量がD/A変換部8のサンプリング周波数である8000で除されることにより、蓄積時間T_mが求められる。そして、予め作成された蓄積時間T_mに対するポーズ継続長T_{de1}のデータに基づいて、蓄積時間T_mに対するポーズ継続長T_{de1}が求められる。

【0180】次の表は、VTRの2倍速再生時における蓄積時間T_mに対するポーズ継続長T_{de1}のデータの一例を示している。

【0181】

【表2】

T _m	0.25～ 0.5 秒	0.5 ～ 0.75秒	0.75～ 1.0 秒	1.0 ～ 1.25秒	1.25～ 1.5 秒	1.5 ～ 1.75秒	1.75～ 2.5 秒	2.5 秒 以上
T _{de1}	0.350 秒	0.300 秒	0.250 秒	0.200 秒	0.150 秒	0.100 秒	0.050 秒	0.025 秒

【0182】以上は、入力信号がアナログ信号の場合について説明したが、入力信号がデジタルデータである場合にもこの発明を適用することができる。たとえば、ICメモリ、磁気ディスク、デジタル通信回線等から、圧縮されたデジタル音声信号が送られてきた場合には、圧縮されたデジタル音声信号が伸長されてPCM音声信号に変換され、得られたPCM音声信号がバッファに一旦格納される。その後、設定された再生速度倍率に応じた速度で、PCM音声データがバッファから読み出されて、図1のフレームメモリ5に送られる。

【0183】以上適応型話速変換処理について詳述したが、次に単純間引き方式による話速変換について具体的なメモリ制御動作について説明する。すなわち、話速変換IC112はメモリ113を次のように制御して話速変換を行う。

【0184】図28は3倍速再生時におけるメモリ制御動作を示しており、音声信号を3倍速のままで書き込み、書き込みと同時に読み出しを始め、音声1倍速再生と同じ速度になるように読み出しを終了する。すなわち、読み出し周期Tの1/3の周期T/3で書き込むようにメモリ制御されるようになっており、従って、図のT0期間が間引かれることになる。同様に、5倍速再生時は、書き込み時間がT/5となり、9倍速再生時はT/9となる。

【0185】図29は逆転5倍速再生時におけるメモリの読み出し/書き込みタイミングを示しており、Tが1倍速周期、Wは書き込み期間、*2はT/5、*はT/6となっている。メモリの書き込みアドレスのタイミングは5倍速であり、書き込みサイクル期間が5倍速正方向再生の5/6となっており、また、T期間6回とした

書き込みサイクルカウンタは、5回で巡回し、アドレス値が0の期間だけ書き込みをさせている。このため図に示すように書き込みと、読み出しとは少しずつずれて巡回し、読み出し5回でもとに戻るのでα点での内容変化は起こらずスムーズに再生音が聞こえる。尚、図に示すように書き込みと読み出しは、dとD、eとE、f1はF1、F2はf2の順で行われる。また、図28及び図29において、破線は書き込みアドレスの歩進状態を示し、実線は読み出しアドレスの歩進状態を示す。

【0186】

【発明の効果】このように本発明のVTRでは、2倍速再生時においては適応型話速変換処理を行い、3倍速再生以上のときは単純間引き処理を行うようにすることにより、話速変換用ICとしてその構成が簡単になるばかりでなく、適切な話速変換処理が自動的に選択されるようになっているのでVTRとしての商品価値を向上せしめることができる。さらに、この発明によれば、2倍速再生時は処理負荷を低減できるとともに、映像と音声のズレを小さくでき、しかも音声信号を蓄積するためのメモリの容量も膨大とならないという利点が得られる。

【図面の簡単な説明】

【図1】話速変換装置の全体的な構成を示すブロック図である。

【図2】話速変換部の構成を示すブロック図である。

【図3】PICOLAを用いて、入力信号を圧縮率2/3で圧縮する方法を示す説明図である。

【図4】波形合成処理部による処理を説明するための説明図である。

【図5】間引き処理部によって行なわれる各種の間引き処理方法を説明するための説明図である。

【図6】話速変換部による処理手順を示すフローチャートである。

【図7】話速変換部による処理手順を示すフローチャートである。

【図8】話速変換部による処理手順の変形例を示し、図7に相当するフローチャートである。

【図9】図6のステップ10の処理と置き換え可能な処理を説明するための説明図である。

【図10】2倍速再生時の入力信号と出力信号との関係を示し、特に無音区間の入力信号が削除される様子を示すタイムチャートである。

【図11】リングメモリ7へのデータ書き込み開始点、リングメモリ7からのデータ読み出し開始点ならびに図10の点A～Dにおけるリングメモリ7の状態を示す模式図である。

【図12】図10の点E～Hにおけるリングメモリ7の状態を示す模式図である。

【図13】2倍速再生時の入力信号と出力信号との関係を示し、特にオーバーフロー直前状態となったときに、入力信号が削除される様子を示すタイムチャートであ

る。

【図14】図13の各点S～Uにおけるリングメモリ7の状態を示す模式図である。

【図15】音声区間と無音区間とを判別するための回路の変形例を示し、図2に相当するブロック図である。

【図16】音声区間と無音区間とを判別するための回路の他の変形例を示し、図2に相当するブロック図である。

【図17】固定フレーム単位で、入力信号を圧縮率2/3で圧縮する方法を示す説明図である。

【図18】図6のステップ9の処理と置き換え可能な処理を説明するための説明図である。

【図19】図6のステップ9の処理として図18の処理を採用した場合に、図6のステップ10の処理と置き換え可能な処理を説明するための説明図である。

【図20】音声区間と無音区間とを判別するための回路のさらに他の変形例を示し、図2に相当するブロック図である。

【図21】定常状態のパワースペクトル、雑音を含まない音声のパワースペクトルおよび音声区間のパワースペクトルを示すグラフである。

【図22】しきい値調整手段およびボーズ継続長調整手段が付加された話速変換部を示すブロック図である。

【図23】本発明を実施したビデオテープレコーダの要部回路ブロック図である。

【図24】図23の回路ブロック図の動作説明のためのフローチャートを示す図である。

【図25】適応型話速変換処理の概念を説明するための図である。

【図26】単純間引き処理による話速変換の概念を説明するための図である。

【図27】逆転再生時における単純間引き処理による話速変換の概念を説明するための図である。

【図28】単純間引き処理を実現するためのメモリ制御方法を説明するための図である。

【図29】逆転再生時における単純間引き処理を実現するためのメモリ制御方法を説明するための図である。

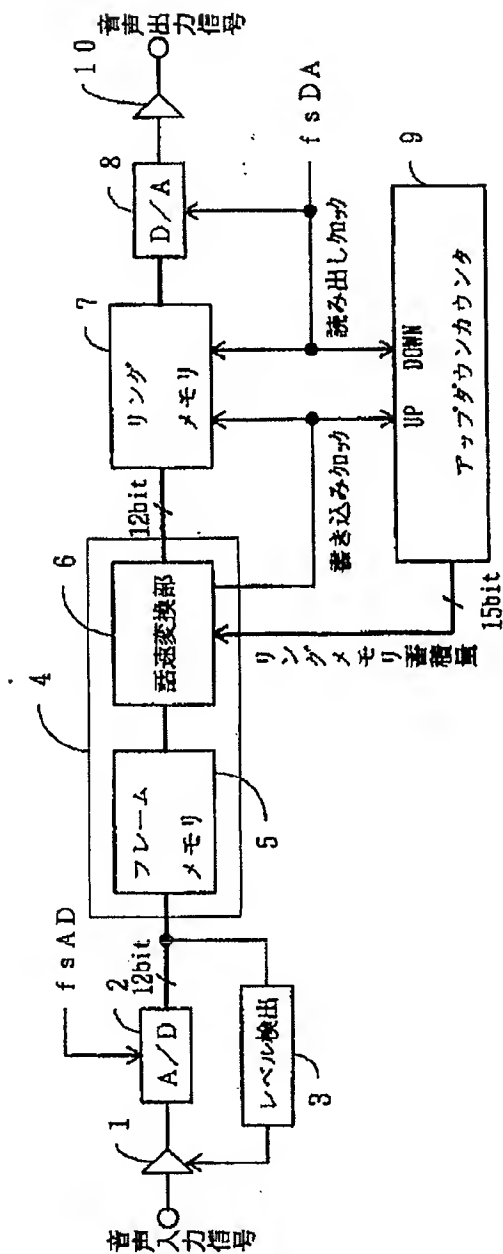
【符号の説明】

- 2 A/D変換部
- 4 DSP
- 5 フレームメモリ
- 6 話速変換部
- 7 リングメモリ
- 8 D/A変換部
- 9 アップダウンカウンタ
- 11 パワー計算部
- 11A 平均振幅計算部
- 11B ビッチ周期検出部
- 11C パワースペクトル計算部
- 12、12A、12B、12C 比較部

- 15 条件分岐部
- 16 リングメモリ蓄積量状態判別部
- 21、25 入力信号削除部
- 23 ピッチ圧縮伸長手段
- 24 間引き処理部

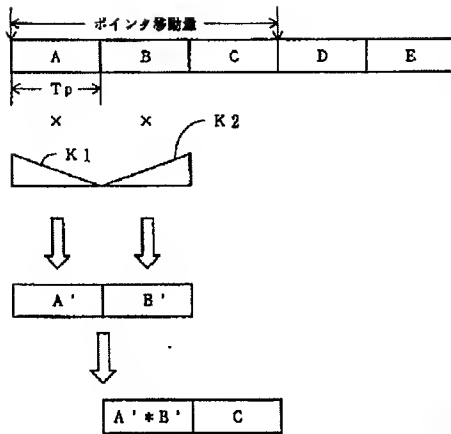
- 51 しきい値調整手段
- 52 ポーズ継続長調整手段
- 112 話速変換IC
- 114 マイコン(マイクロコンピュータ)

【図 1】

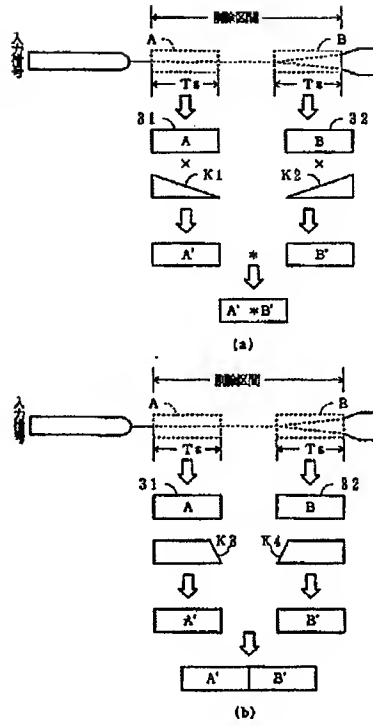


[illegible]

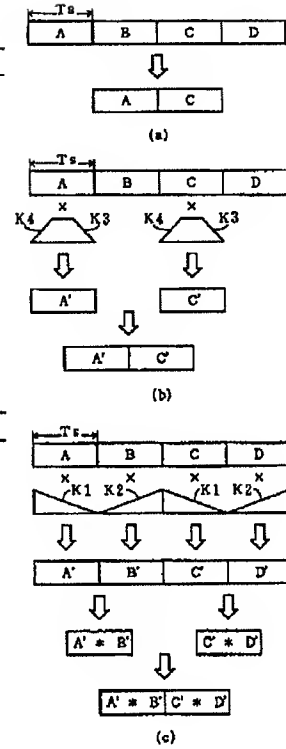
【図3】



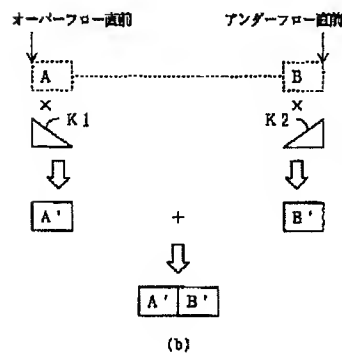
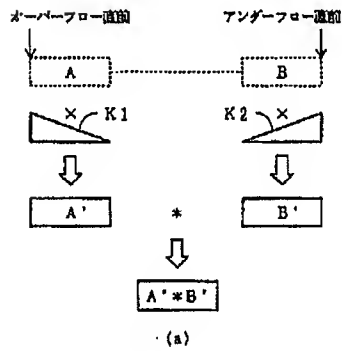
【図4】



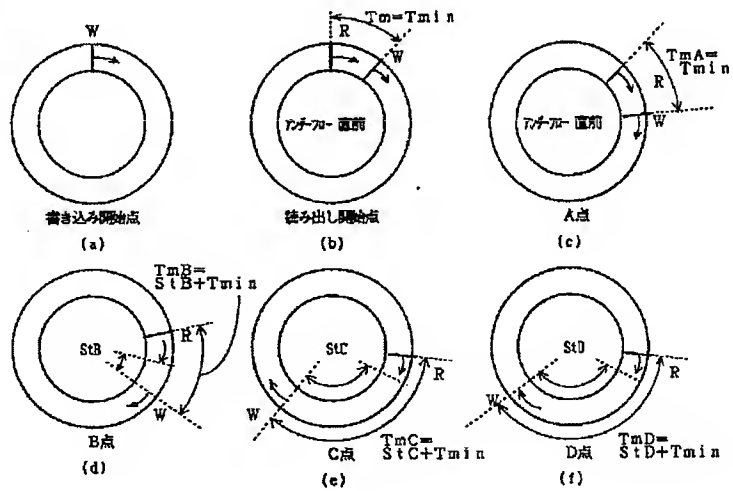
【図5】



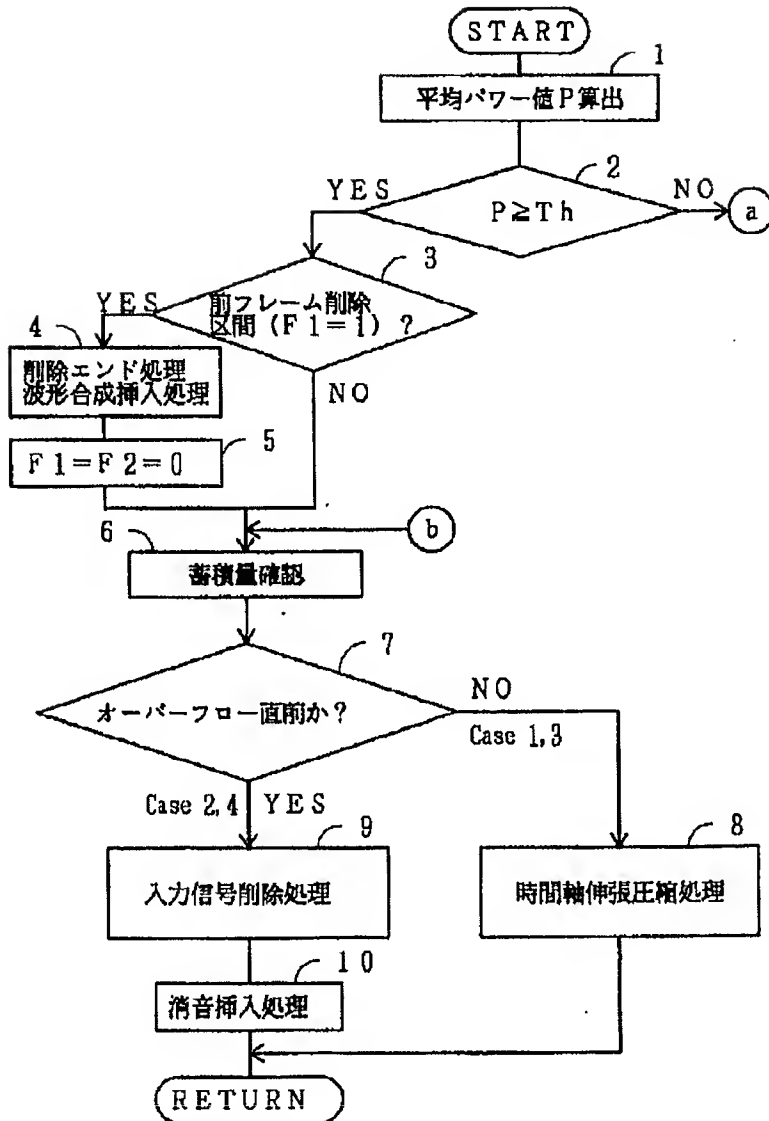
【図9】



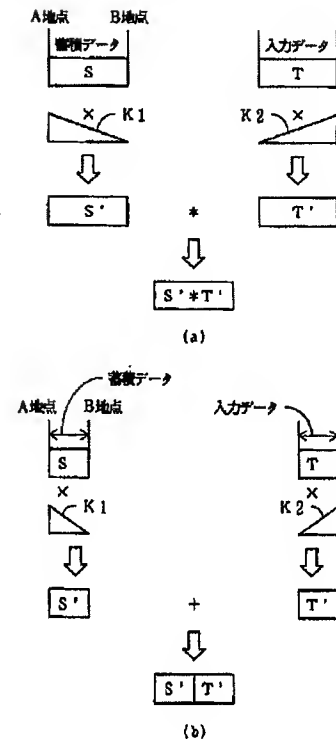
【図11】



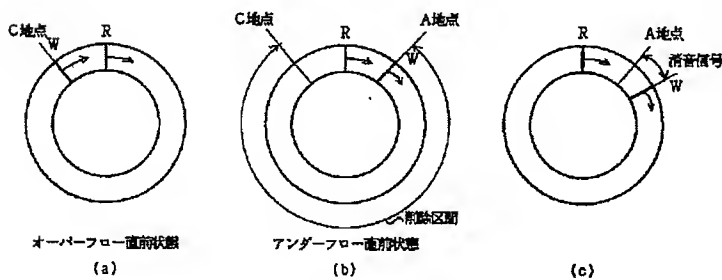
【図6】



【図19】



【図18】

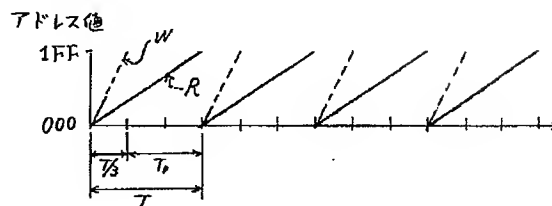
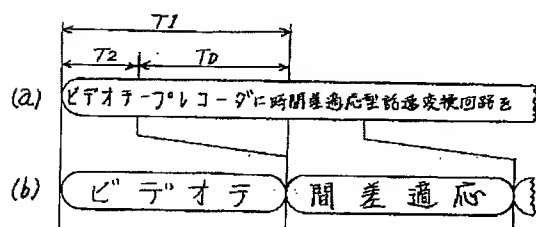


```

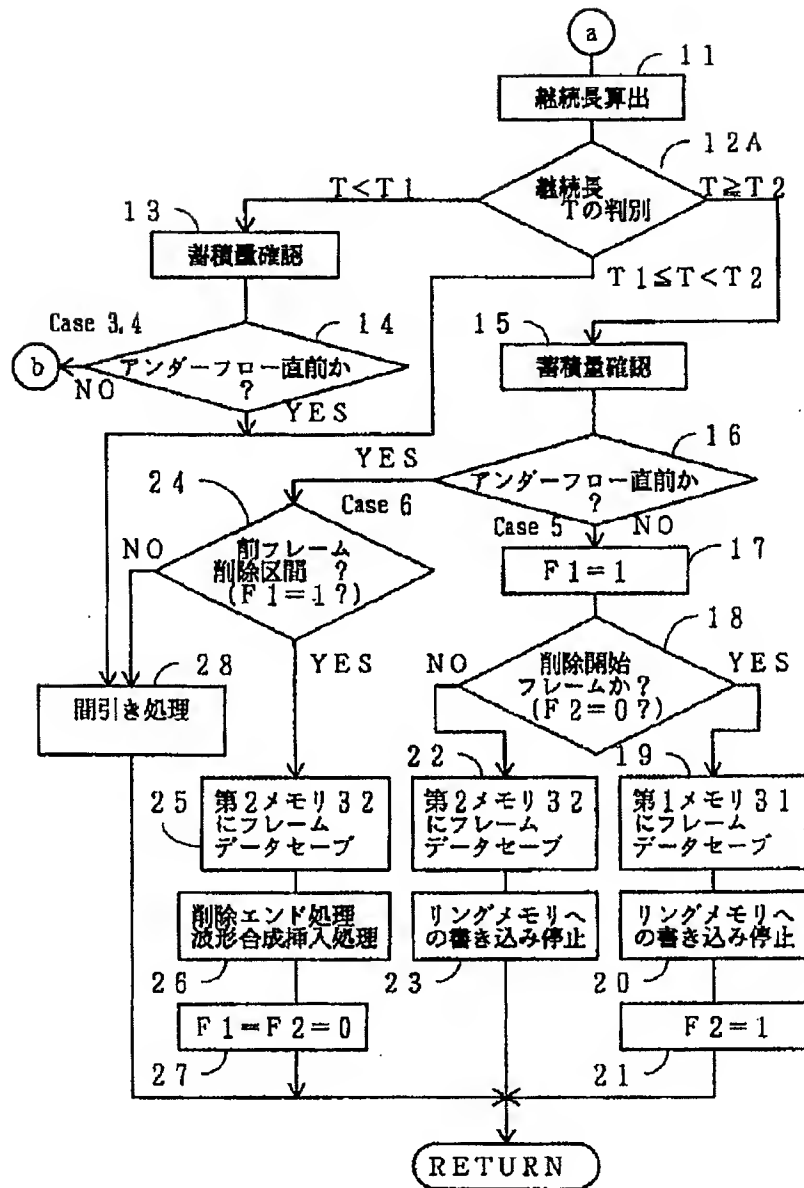
graph TD
    Start((a)) --> 11[継続長算出]
    11 --> 12{継続長が設定値  
Tdel以上か}
    12 -- NO --> 13[蓄積量確認]
    12 -- YES --> 15[蓄積量確認]
    13 --> 14{アンダーフロー直前か}
    14 -- Case 3,4 NO --> b((b))
    14 -- YES --> 24{前フレーム削除区間  
(F1=1?)}
    24 -- NO --> 28[間引き処理]
    24 -- YES --> 25[第2メモリ32  
にフレーム  
データセーブ]
    25 --> 26[削除エンド処理  
波形合成挿入処理]
    26 --> 27[F1=F2=0]
    27 --> 21(( ))
    15 --> 16{アンダーフロー直前か}
    16 -- Case 5 NO --> 17[F1=1]
    16 -- YES --> 18{削除  
開始フレームか?  
(F2=0?)}
    17 --> 22[第2メモリ32  
にフレーム  
データセーブ]
    18 -- NO --> 22
    18 -- YES --> 19[第1メモリ31  
にフレーム  
データセーブ]
    19 --> 20[リングメモリへの  
書き込み停止]
    20 --> 21[F2=1]
    21 --> 21
    21 --> 21
    21 --> RETURN([RETURN])

```

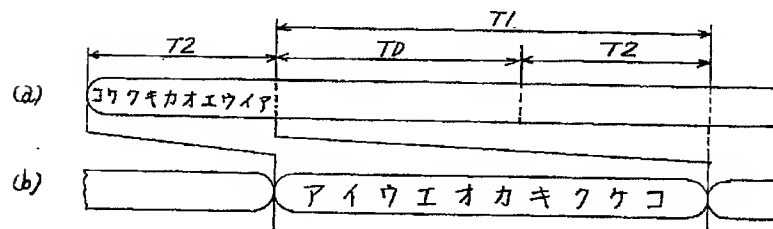
【图 28】



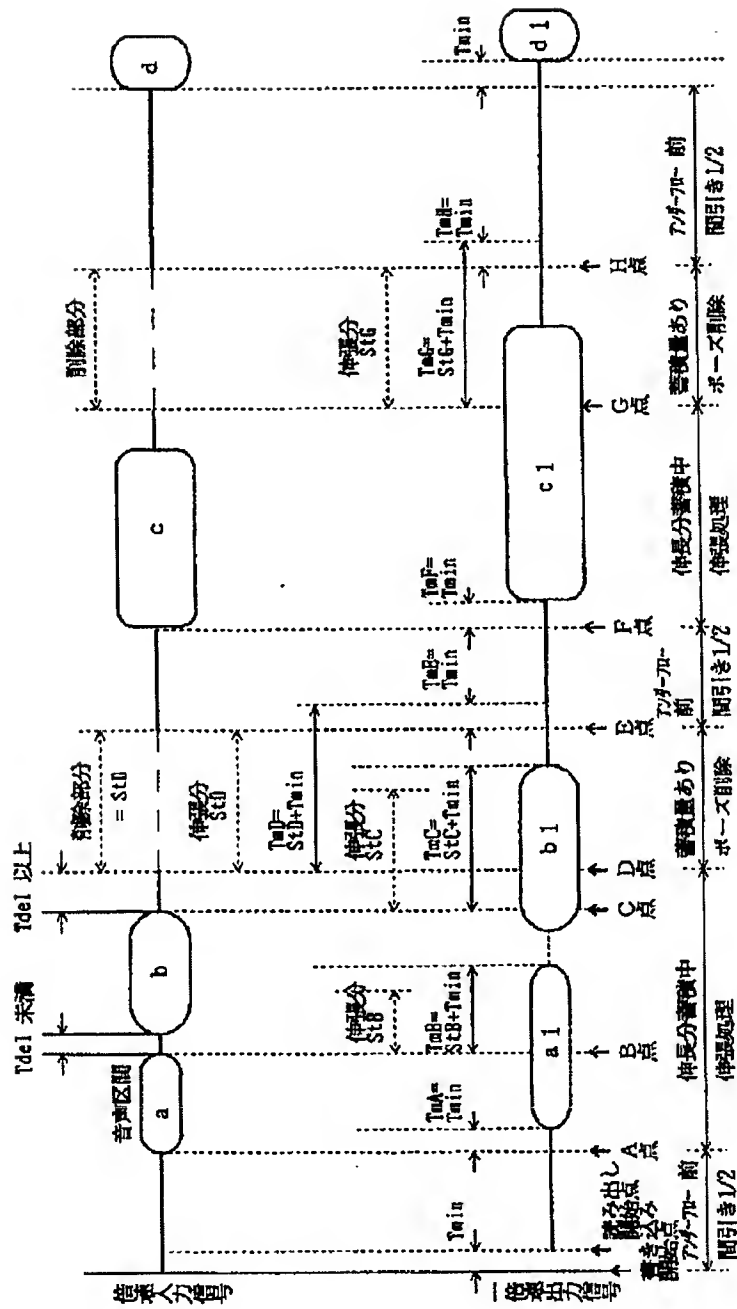
【図8】



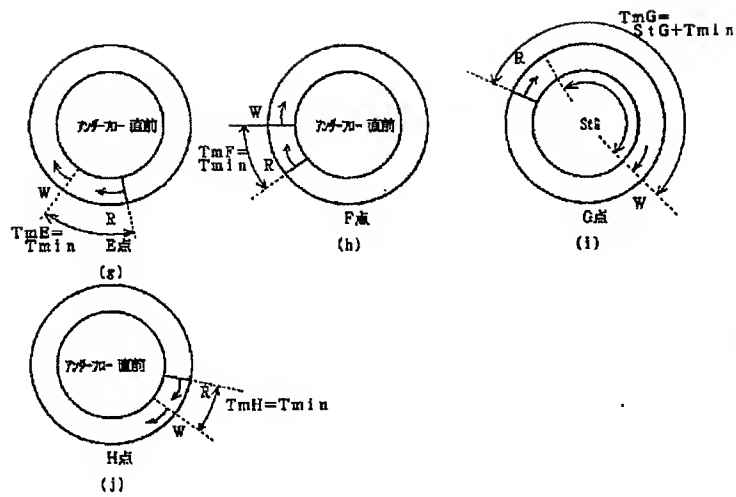
【図27】



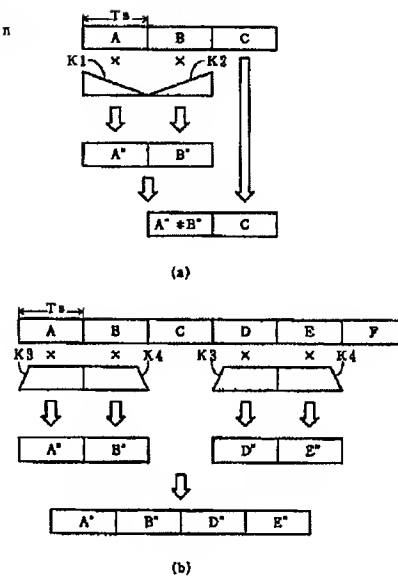
【図10】



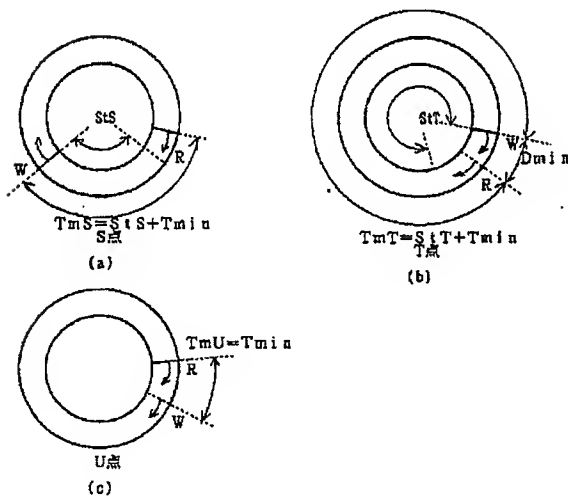
【図12】



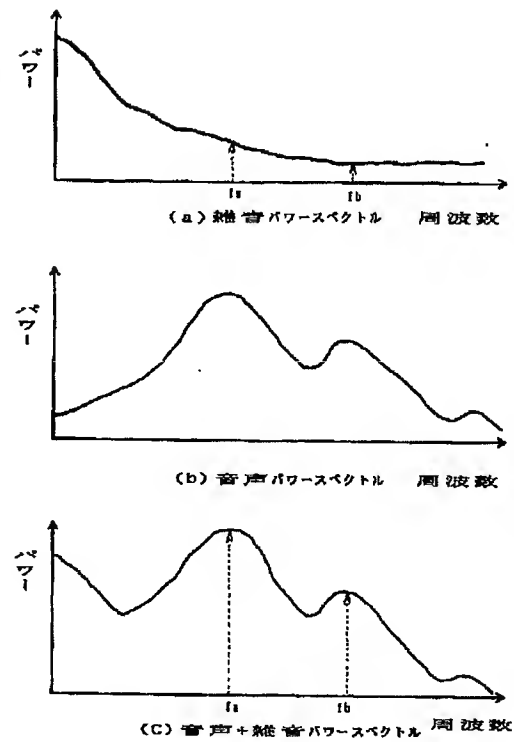
【図17】



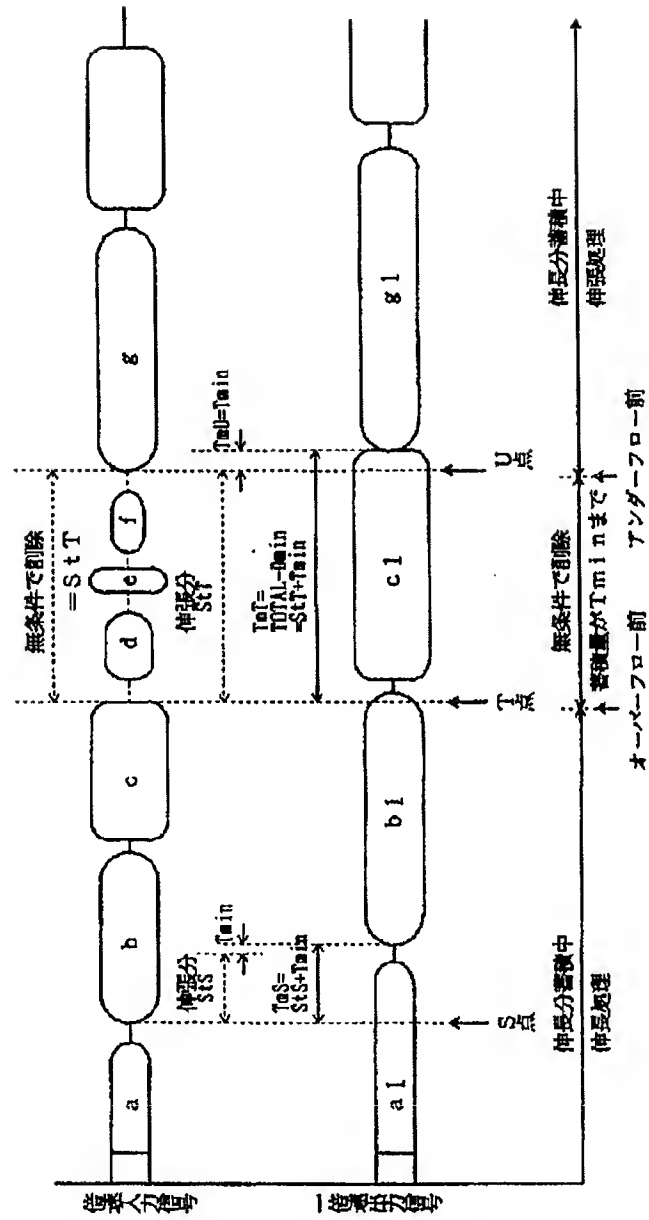
【図14】



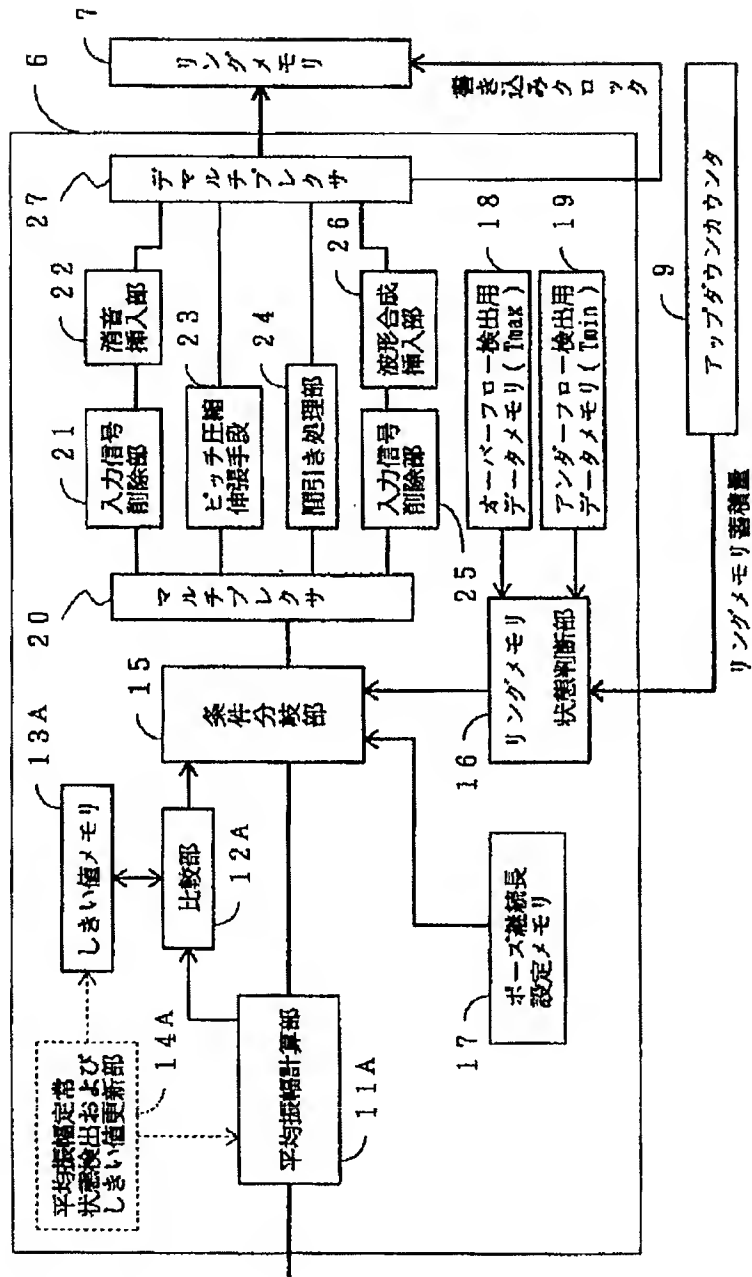
【図21】



【図13】

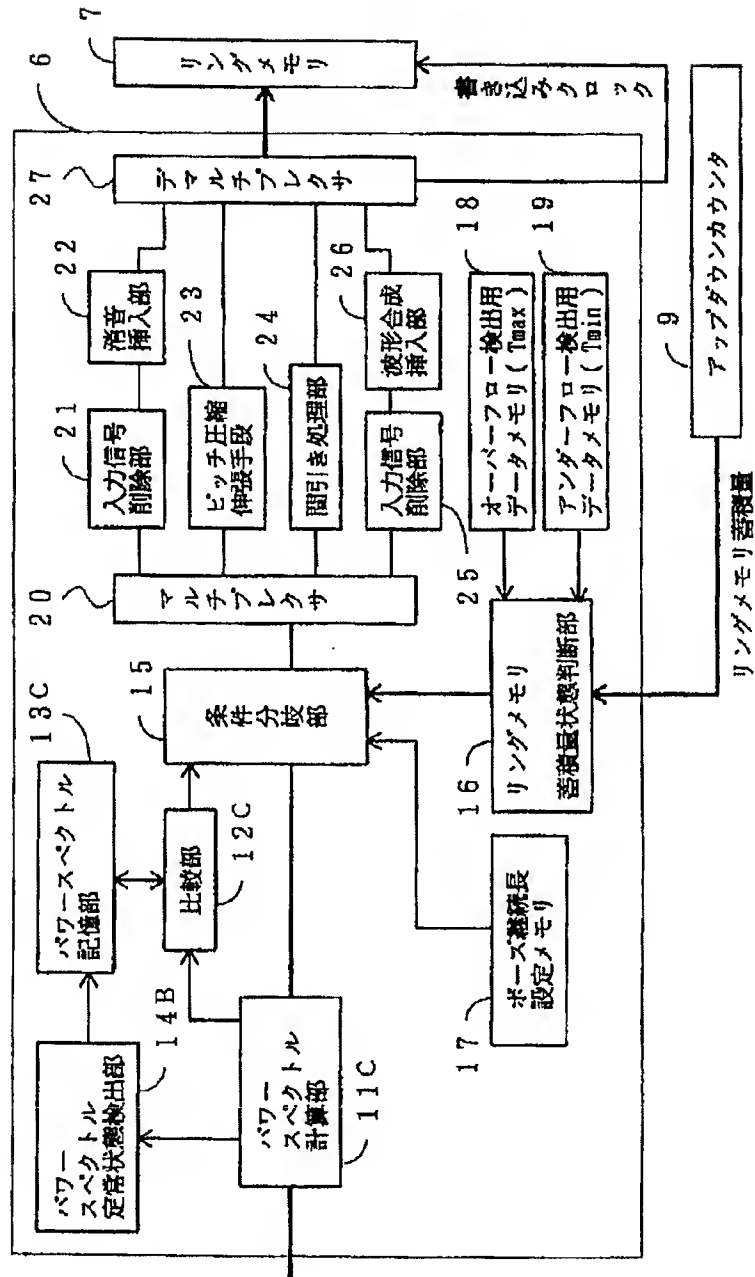


【図15】

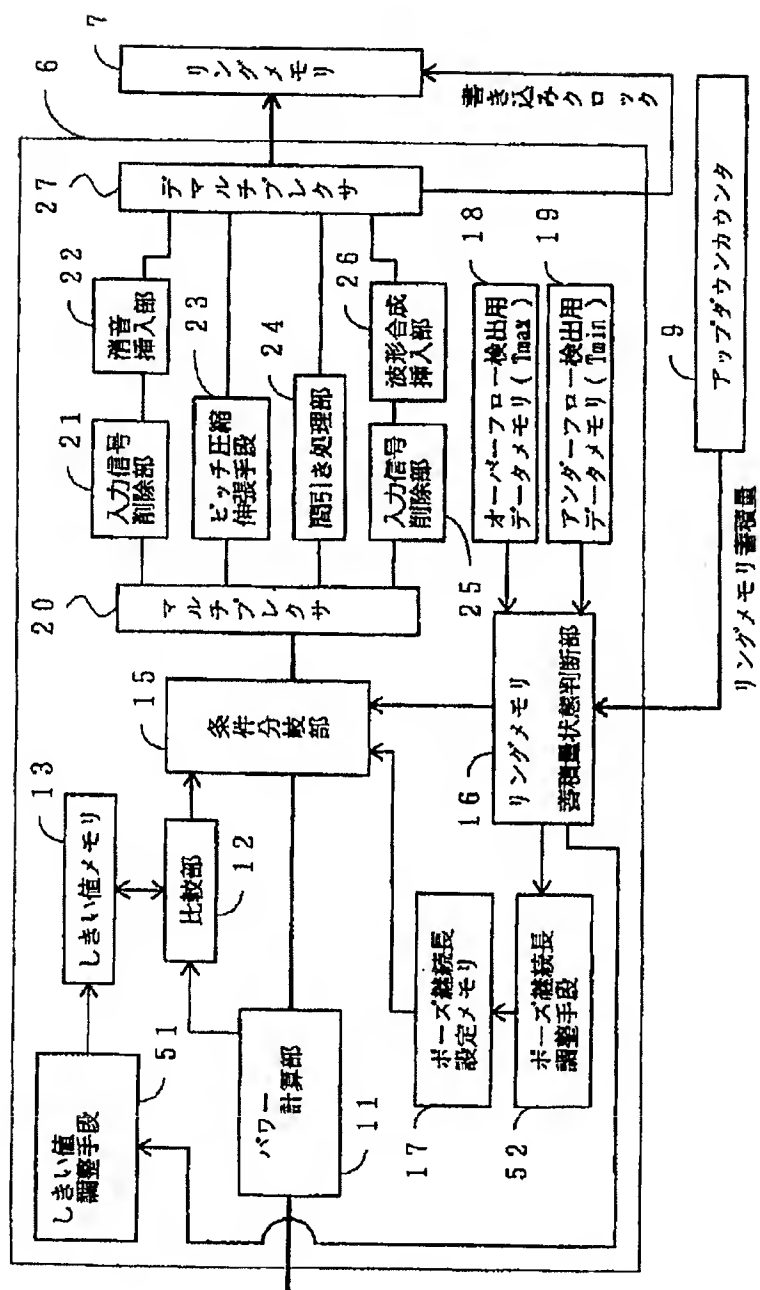


[illegible]

【図20】



【图22】



(a)

話速変換を

用いた

VTRです

無音区間

T

(b)

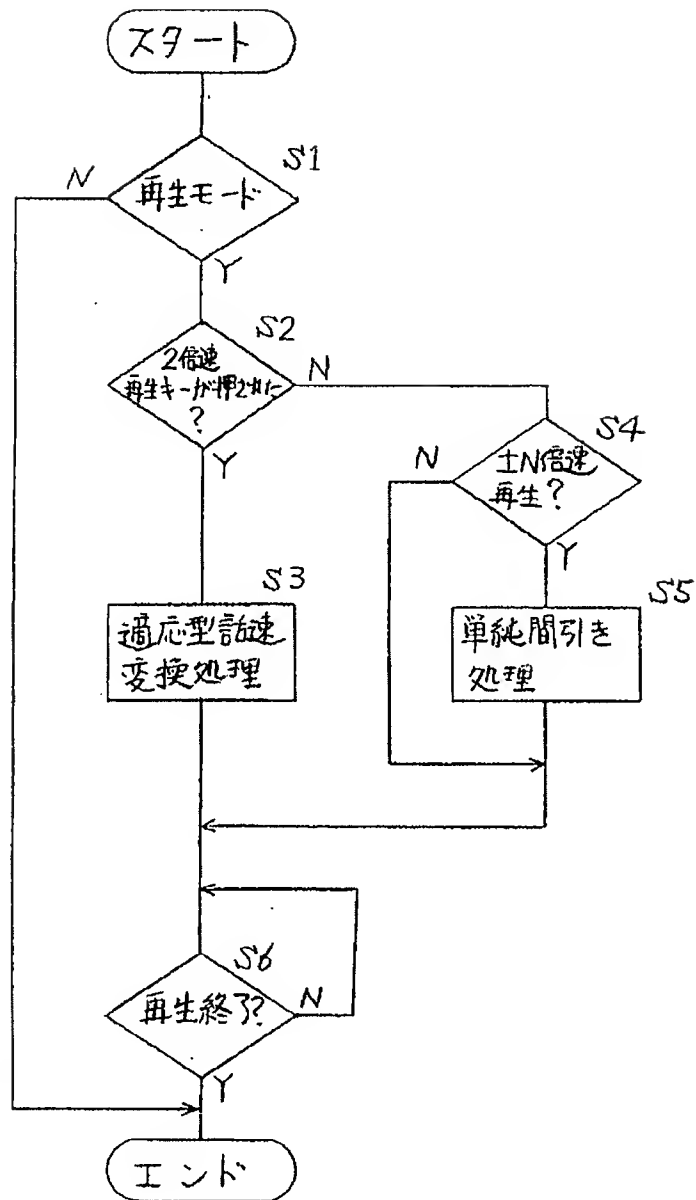
話速変換を

用いた

VTRです

$T/2$

【図24】



フロントページの続き

(51)Int. Cl.⁶
H04N 5/93

識別記号 序内整理番号

FI
H04N 5/93

技術表示箇所

G